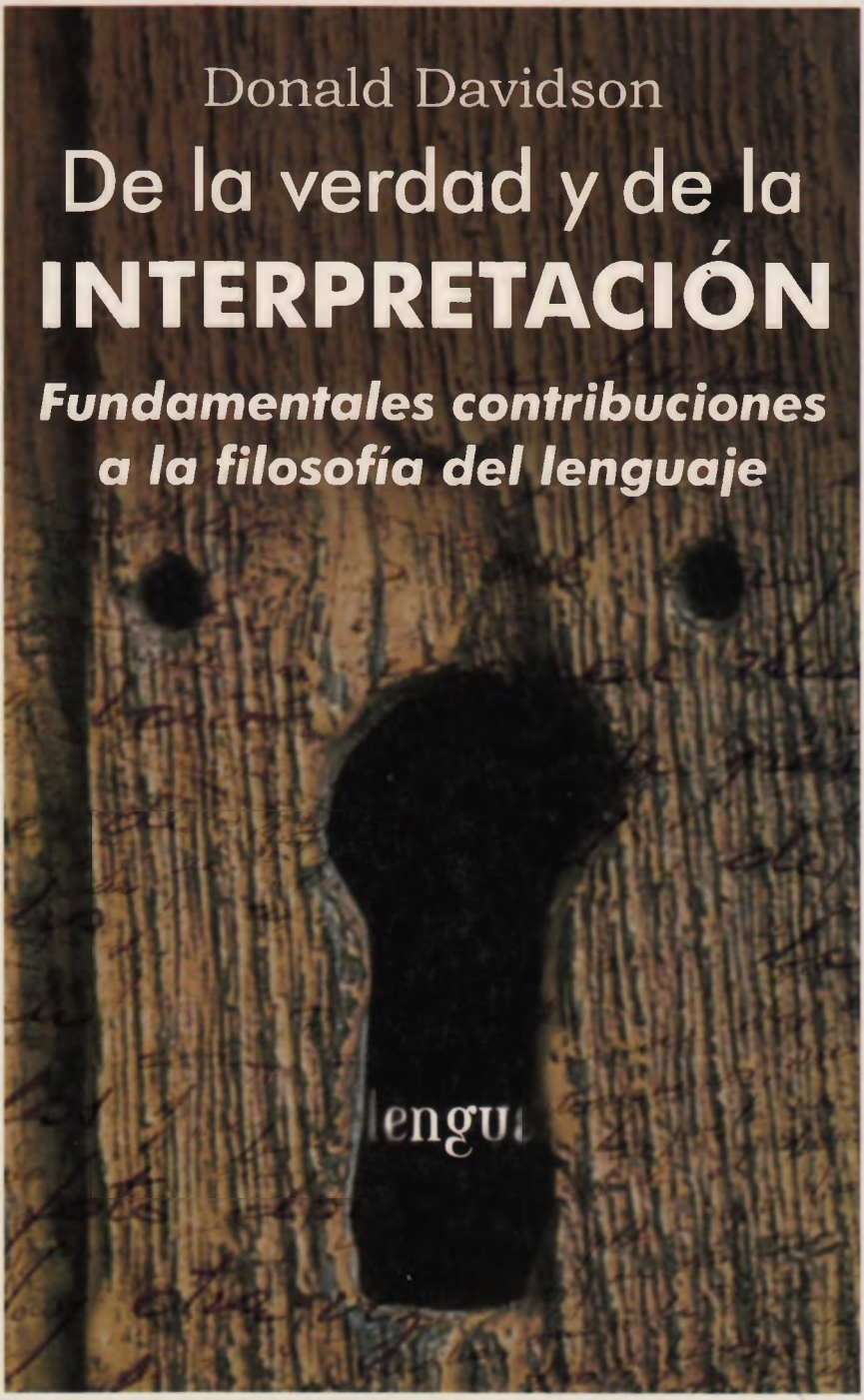
Filosofía/Lingüística

Donald Davidson

De la verdad y de la INTERPRETACIÓN

Fundamentales contribuciones a la filosofía del lenguaje



gedisa
editorial

QUIMERA LIBROS
NUEVA DE LYON 045 • LOCAL 8
FONO 232 8266 • PROVIDENCIA
E-mail: libreria@quimeralibros.cl

Donald Davidson

DE LA VERDAD Y DE LA INTERPRETACION

Serie
CLA•DE•MA
Filosofía / Lingüística

DE LA VERDAD Y DE LA INTERPRETACION

Fundamentales contribuciones a la filosofía del lenguaje

Donald Davidson

gedisa
editorial

Título del original en inglés:
Inquiries into Truth and Interpretation
Originally published in English by Oxford University Press
© in this Collection by Donald Davidson, 1984

Traducción: Guido Filippi

Ilustración de la cubierta: Edgardo Carosia

Primera edición, Barcelona, 1990
Primera reimpresión, Barcelona, 1995
Segunda reimpresión, Barcelona, 2001

CREATIVE COMMONS

Derechos reservados para todas las ediciones en castellano

© Editorial Gedisa, S.A.
Paseo Bonanova, 9 1º-1ª
08022 Barcelona (España)
Tel. 93 253 09 04
Fax 93 253 09 05
Correo electrónico: gedisa@gedisa.com
http://www.gedisa.com

ISBN: 84–7432–342-8
Depósito legal: B. 30204 - 2001

Impreso por: Carvigraf
Cot, 31, Ripollet

Impreso en España
Printed in Spain

Para W. V. Quine

INDICE

VOCABULARIO UTILIZADO 11

PROCEDENCIA DE LOS ENSAYOS Y RECONOCIMIENTOS 13

INTRODUCCION 17

Verdad y significado

1. Teorías del significado y lenguajes aprendibles (1965) 27
2. Verdad y significado (1967) 39
3. Fiel a los hechos (1969) 57
4. Semántica para lenguajes naturales (1970) 73
5. En defensa de la Convención T (1973) 82

Aplicaciones

6. La cita (1979) 95
7. Al decir *that* (1978) 108
8. Modos y ejecuciones (1979) 123

Interpretación radical

9. Interpretción radical (1973) 137
10. La creencia y el fundamento del significado (1974) 151
11. Pensamiento y habla (1975) 164
12. Respuesta a Foster (1976) 179

Lenguaje y realidad

13. De la idea misma de un esquema conceptual (1974) 189
14. El método de la verdad en metafísica (1977) 204
15. Realidad sin referencia (1977) 219
16. La inescrutabilidad de la referencia (1979) 229

Límites de lo literal

17. Qué significan las metáforas (1978) 245
18. Comunicación y convención (1982) 263

REFERENCIAS BIBLIOGRAFICAS 277

INDICE TEMATICO 283

Vocabulario utilizado*

Background language = lenguaje de encuadre
Background theory = teoría de encuadre
Building-block theory = teoría del bloque constructivo
Counterfactural = contrafáctico
Embedded sentence = oración anidada
Events = sucesos
Facts = hechos
Indexical = deíctico
Malapropism = malapropismo
Map (to) = mapear
Mass-term = término masa
Meaningfulness = significatividad
Miniature theory = teoría mínima
Mood-setter = fijador de modo
Object-language = lenguaje objeto
Ostend = ostentar
Ostension = ostensión
Performance = realización
Performative = realizativo
Phrase-structure gramma = gramática de estructura de frase
Picturing = ilustración
Posit, posited = supuesto
Quotation-mark name = nombre-comilla
Quotational theory of = teoría citativa de
Sense-data = sense data (o datos sensoriales
Sensory promptings = incitaciones sensoriales
Significant = significativo
Spelling theory of = teoría del deletreo de
Stimulus meaning = significado-estímulo
Surface irritations = irritaciones de superficie
Token = instancia particular (o ejemplar)
T-sentences = oraciones T
Utterance = emisión (o proferencia)
wh-questions = preguntas wh
Yes-no interrogatives = interrogativas sí-no

* Este pequeño glosario contiene los términos básicos empleados en este libro; han sido sugeridos por el Dr. Davidson y la lista definitiva fue confeccionada por el traductor con la colaboración generosa de la prof. Olbeth Rossi. [E.]

Procedencia de los ensayos* y reconocimientos

El Ensayo 1, "Teorías del significado y lenguajes aprendibles", se leyó en el Congreso Internacional de Lógica, Metodología y Filosofía de la Ciencia en la Universidad Hebrea de Jerusalén, en 1964. Se publicó posteriormente en las Actas de dicho congreso en un volumen a cargo de Yehoshua Bar-Hillel publicado por North-Holland Publishing Company, Amsterdam, 1965. Se reproduce aquí con autorización de los editores.

Una versión anterior del Ensayo 2, "Verdad y significado", se leyó en la reunión de la Eastern Division de la American Philosophical Association, en diciembre de 1966. El tema central tiene su origen en un trabajo presentado a la Pacific Division en 1962. El presente ensayo debe mucho a John Wallace, con quien discutí estos asuntos desde 1962 en adelante. Mis investigaciones fueron auspiciadas por la National Sciense Foundation. El trabajo se publicó por primera vez en *Synthèse*, 17 (1967), 304-23, copyright 1967 por D. Reidel Publishing Company, Dordrecht-Holland, reproducido con autorización de D. Reidel Publishing Company.

El Ensayo 3, "Fiel a los hechos", se presentó por primera vez en un simposio sobre La Verdad en diciembre de 1969, en una reunión de la Eastern Division de la American Philosophical Association. La otra ponencia estuvo a cargo de James F. Thomson. El trabajo se publicó por primera vez en el *Journal of Philosophy*, 66 (1969), 748-64, y se reproduce aquí con autorización de los editores.

"Semántica para lenguajes naturales", el Ensayo 4, se leyó en un simposio organizado por la compañía Olivetti en honor a su fundador, que tuvo lugar en Milán en octubre de 1968. Las actas se publicaron en *Linguaggi nella Società e nella Tecnica*, Edizioni di Comunità, Milán, 1970.

El Ensayo 5, "En defensa de la Convención T", se leyó en una conferencia sobre Semántica Alternativa efectuada en la Temple University en diciembre de 1970 y se publicó en *Truth, Syntax and Modality* (lamentablemente, se rechazó el título *Truth Valued* sugerido por Dana Scott). La preparación del libro estuvo a cargo de Hughes Leblanc y lo publicó North-Holland Publishing Company, 1973. Se reproduce aquí con autorización de los editores.

* El autor denomina a cada uno de los 18 capítulos de que se compone esta obra "ensayo". [T.]

El siguiente Ensayo, el número 6, "La cita", se publicó en un número especial de *Theory and Decision* sobre Teoría del Lenguaje compilado por H. L. Berghel [*Theory and Decision*, 11 (1979), 27-40], copyright 1979 de D. Reidel Publishing Company, Dordrecht-Holland. Reproducido con autorización de D. Reidel Publishing Company.

"Al decir *That*", el Ensayo 7, se publicó en un número doble de *Synthèse* dedicado a la obra de W. V. Quine [*Synthèse*, 19 (1968-9) 130-46]. Se publicó más tarde en Words and Objections, *Essays on the Work of W. V. Quine*, a cargo de D. Davidson y J. Hintikka, D. Reidel, 1969, págs. 158-74 (edición revisada, 1975), copyright 1969 por D. Reidel Publishing Company, Dordrecht, Holanda. Reproducido con autorización de D. Reidel Publishing Company. La respuesta de Quine a este ensayo está en las páginas 333-5 de *Words and Objections*.

El Ensayo 8, "Modos y ejecuciones", se leyó en el segundo Encuentro Filosófico de Jerusalén, llevado a cabo en Israel en abril de 1976, y fue comentado por W. V. Quine. Se publicó en *Meaning and Use*, compilado por A. Margalit, D. Reidel Publishing Company, Dordrecht, Holanda, 1979.

"Interpretación radical", el Ensayo 9, se leyó en un coloquio sobre Problemas Filosóficos del Lenguaje, en Biel, Suiza, en mayo de 1973, y una segunda versión en una conferencia sobre Lenguaje y Significado en Cumberland Lodge, Greant Park, Windsor, en noviembre de ese año. Se publicó en Cumberland Lodge, Great Park, Windsor, en noviembre de ese año. Se publicó en *Dialectica*, 27 (1973), 313-28, y se reproduce aquí con el permiso del compilador H. Lauener.

El Ensayo 10 sobre "La creencia y el fundamento del significado" se preparó para una conferencia sobre Teoría de Lenguaje, Intencionalidad y Traducción que se realizó en la University of Connecticut en marzo de 1973, y se publicó en un número doble de *Synthèse* a cargo de J. G. Troyer y S. C. Wheeler III [*Synthèse*, 27 (1974), 309-23]. Este número también incluye valiosos comentarios de W. V. Quine (325-9) y David Lewis (331-44) y las respuestas (345-9). La respuesta de David Lewis lleva por título "Interpretación radical", copyright 1974 de D. Reidel Publishing Company, Dordrecht, Holanda. Reproducido con autorización de D. Reidel Publishing Company.

El Ensayo 11, "Pensamiento y habla", corresponde a una conferencia que di en 1974 en el Wolfson College, publicada en *Mind and Language* a cargo de Samuel Guttenplan, copyright Oxford University Press, 1975. Se reproduce aquí con autorización de Oxford University Press en nombre del Wolfson College.

En junio de 1974 John Foster leyó un trabajo, "Teoría del significado y verdad", ante la Oxford Philosophical Society, y el Ensayo 12, "Respuesta a Foster". Es mi respuesta a las partes de su trabajo que tenían relación con mi obra. El trabajo de Foster y mi respuesta se publicaron en *Truth and Meaning: Essays in Semantics*, a cargo de Gareth Evans y John McDowell, copyright Oxford University Press, 1976.

El Ensayo 13, "De la idea misma de un esquema conceptual", tardó en alcanzar su forma actual. La sexta y última de mis conferencias so-

bre John Locke, titulada "Invariantes de traducción", que ofrecí en 1970 en Trinity Term, Oxford, fue un primer borrador. En enero del año siguiente di dos conferencias sobre Esquemas Conceptuales Alternativos en la University of London que contenían buena parte de lo que hoy es el ensayo. Destilé el material casi hasta su forma final para mi discurso presidencial en un encuentro de la Eastern Division de la American Philosophical Association en Atlanta, el 28 de diciembre de 1973. Más tarde, antes de su publicación, di una charla afín, "The Third Dogma of Empiricism", en la Philosophical Society de Oxford. W. V. Quine abrió la discusión, y sus reflexiones me ayudaron a redactar el borrador final. Algunas derivaciones posteriores de aquella discusión se encontraron en "On the Very Idea of a Third Dogma", de Quine. Mi trabajo es publicó en los *Proceedings and Addresses of the American Philosophical Association*, 47 (1974), y se reproduce aquí con autorización de la Association.

El Ensayo 14, "El método de la verdad en metafísica", se publicó primeramente en *Midwest Studies in Philosophy, 2: Studies in the Philosophy on Languaje*, a cargo de P. A. French, T. E. Uehling (Jr.) y H. K. Wettstein, The University of Minnesota, Morris, 1977. Me resultaron de gran utilidad los comentarios que sobre versiones previas efectuaron Gilbert Harman y W. V. Quine.

El Ensayo 15, "Realidad sin referencia", se presentó por primera vez en una versión diferente en un Taller de Semántica y Lingüística en la University of Western Ontario, London, Ontario, en abril de 1972. Críticas y sugerencias subsiguientes modificaron el trabajo a lo largo de los años. Se publicó por primera vez en *Dialectica*, 31 (1977), 247-53, y se reproduce aquí con autorización de los editores.

"La inescrutabilidad de la referencia", el Ensayo 16, se escribió para un número especial de *The Southwestern Journal of Philosophy* dedicado a la obra de W. V. Quine, pero no lo terminé a tiempo y el trabajo apareció en un número posterior. *The Southwestern Journal of Philosophy,* 10 (1979), 7-19. Se reproduce aquí con autorización del *Journal.* Quine respondió a mi trabajo y a otros en "Replies to the Eleven Enssays", *Philosophical Topics*, 11 (1981), 242-3.

El Ensayo 17, "Qué significan las metáforas", se leyó en una conferencia sobre La Metáfora en la University of Chicago en febrero de 1978. Se publicó por primera vez en *Critical Inquiry*, 5 (1978), 31-47, copyright 1978 de Donald Davidson. El extracto de "The Hippopotamus" en la pág. 256 pertenece a *Collect Poems 1909-1962,* de T. S. Eliot, copyright 1936, de Faber and Faber and Harcourt Brace Jovanovich Inc., coypright 1963, 1964 de T. S. Eliot. Se reproduce aquí con permiso de los editores. En un número posterior del mismo periódico, Nelson Goodman y Max Black respondieron a mi trabajo. Ambos evidenciaron disentir con mucho de lo que yo había escrito. [Nelson Goodman, "Metaphoras Moonlighting", *Critical Inquiry*, 6 (1979), 125-30 y Max Black, *"How Metaphore Work: A Replay to Donald Davison"*, *Critical Inquiry*, 6 (1979), 131-43].

El último ensayo, el Ensayo 18, sobre "Comunicación y conven-

ción", se leyó en el primer Encuentro de Campinas sobre la Filosofía del Lenguaje en la Universidad Estadual de Campinas en agosto de 1981. Está prevista su publicación en las actas del Encuentro, *Dialogue: an Interdisciplinary Approach*, a cargo de Marcelo Dascal, John Benjamins, Amsterdam.

Introducción

¿Qué es para las palabras significar lo que significan? En los ensayos aquí reunidos exploro la idea de que tendríamos una respuesta a esta cuestión si supiéramos cómo construir una teoría que satisficiera dos requerimientos: proporcionar una interpretación de todas las emisiones, reales y potenciales, de un hablante o grupo de hablantes; y ser verificable sin conocimiento de las actitudes proposicionales detalladas del hablante. La primera condición reconoce la naturaleza holística de la comprensión lingüística. La segunda condición tiende a prevenir la introducción subrepticia en los fundamentos de la teoría de conceptos excesivamente emparentados con el concepto de significado. Una teoría que no satisfaga ambas condiciones no responde a nuestra pregunta inicial en una forma filosóficamente instructiva.

Los primeros cinco ensayos se ocupan principalmente de sondear qué tipo de teoría podría satisfacer la primera condición.

El Ensayo 1, "Teorías del significado y lenguajes aprendibles", insiste en que una teoría satisfactoria, si es que se pretende su utilidad para una criatura con poderes finitos, debe descubrir un vocabulario básico finito en los fenómenos verbales a ser interpretados. Si esto es así, no se podrá escapar a la necesidad de tratar los rasgos semánticos de la infinidad de potenciales oraciones en base a los rasgos semánticos de los ítems que integran un vocabulario finito. Resulta ser que una cantidad de teorías que nos son familiares no cumplen esta condición: el análisis de Frege de contextos oblicuos, la lógica de sentido y denotación de Church, el tratamiento informal de la cita por parte de Tarski son ejemplos de ello. Las teorías usuales de modificación adverbial bien podrían agregarse a la lista.

El Ensayo 2, "Verdad y significado", sostiene que una teoría de la verdad según los lineamientos de las definiciones de verdad de Tarski, pero modificada en varios aspectos para su aplicación a un lenguaje natural, bastaría para que un intérprete pudiera operar. Dichas teorías tienen claras virtudes. No utilizan los significados como entidades; no se introduce objetos para corresponder a predicados u oraciones; y a partir de un conjunto finito de axiomas es posible probar, para cada oración del lenguaje a ser interpretado, un teorema que enuncia condiciones de verdad para dicha oración. Además, la comprobación de tal teorema representa un análisis del grado en que la verdad o falsedad de la oración depende de cómo está compuesta a partir de elementos tomados del vo-

cabulario básico. Si dichas teorías satisfacen realmente las dos condiciones descritas en el primer párrafo, podemos tomar en serio la palabra "teoría" en "teoría del significado".

Se han hecho oír muchas objeciones a la pretensión de que las teorías de la verdad pueden funcionar como teorías del significado. En otros ensayos de este libro he intentado enfrentar o desviar algunas de las objeciones. Pero resulte o no aceptable esa pretensión, algunos de los argumentos a su favor presentes en "Verdad y Significado" están viciados. El lector observará que cambié de terreno más de una vez en mi intento por mejorar o clarificar esta tesis central. Algo que sólo comencé a vislumbrar gradualmente fue que mientras Tarski procuraba analizar el concepto de verdad apelando (en la Convención T) al concepto de significado (con la forma de igualdad de significado, o traducción), yo tenía en mente lo contrario. Consideraba la verdad como concepto primitivo central, y esperaba llegar al significado mediante la búsqueda en detalle de la estructura de la verdad. Desde luego, éstas son observaciones *acerca* de las teorías de la verdad y no observaciones a ser encontradas en ellas.

Había algo más que comenzaba a percibir lentamente: puesto que yo estaba tratando a las teorías de la verdad como teorías empíricas, los axiomas y teoremas tenían que ser considerados como leyes. Así, un teorema como " 'Schnee ist weiss' es verdadero en boca de un hablante alemán si y sólo si la nieve es blanca" debe no solamente considerarse verdadero, sino también capaz de resistir la acusación de contrafáctico. Ciertamente, dado que la evidencia de esta ley, si lo es, depende en última instancia de ciertas relaciones causales entre los hablantes y el mundo, uno puede decir que no es un accidente que "Schnee ist weiss" sea verdadero si y sólo si la nieve es blanca; es la blancura de la nieve la que *hace* a "Schnee ist weiss" verdadero. Hasta qué punto esto es una concesión a la intensionalidad depende, supongo, del propio análisis del concepto de ley. Parece claro que, sea cual fuere la concesión, ésta debe hacerse para cualquier ciencia empírica. Estas cuestiones se discuten en el Ensayo 12.

El Ensayo 3, "Fiel a los hechos", se cuestiona si una teoría de la verdad al estilo de Tarski debería considerarse una teoría de correspondencia. Dichas teorías, a diferencia de la mayoría de las teorías de correspondencia, no explican la verdad encontrando entidades tales como los hechos a los cuales deben corresponder las oraciones verdaderas. Y hay buenas razones, que se pueden rastrear hasta Frege, para descartar los hechos como entidades capaces de cumplir este papel. Por otro lado, las teorías de la verdad del tipo considerado aquí requieren efectivamente la caracterización de una relación entre entidades y expresiones ("satisfacción"). No es fácil ver cómo una vía satisfactoria hacia la verdad puede escapar a este paso si el lenguaje que la teoría trata contiene los recursos cuantificacionales habituales.

"Semántica para lenguajes naturales", el Ensayo 4, argumenta que las teorías de la verdad podrían proporcionar una semántica formal para que los lenguajes naturales coincidieran con el tipo de sintaxis formal

que de Chomski en adelante han impulsado los lingüistas. Cuando se escribió este ensayo, se pensaba que las estructuras profundas de la sintaxis eran los vehículos de la interpretación semántica. El Ensayo 4 sugirió que la estructura profunda de una oración debía corresponder a la forma lógica que una teoría de la verdad asignaba a dicha oración.

La Convención T de Tarski, defendida en el Ensayo 5, es un instrumento informal pero poderoso para la puesta a prueba de teorías de la verdad, a diferencia de lo que uno mismo veía antes en este concepto. En su aplicación más directa, el test simplemente nos incita a reconocer el aspecto no citativo de los predicados de verdad; oraciones como: " 'La nieve es blanca' es verdadero en español si y sólo si la nieve es blanca" son trivialmente verdaderas. Puesto que la totalidad de tales oraciones determina de manera singular la extensión de un predicado de verdad para el español, una teoría que vincule a todas las oraciones de ese tipo debe ser extensionalmente correcta. Los críticos han caído a menudo en el error de pensar que, dado que los teoremas que demuestran que una teoría es correcta son triviales, la teoría o el concepto de verdad que ella caracteriza deben también ser triviales.

Una teoría de la verdad servirá para interpretar a un hablante sólo si la teoría está en condiciones de tener en cuenta todos los recursos lingüísticos del hablante. Pero, ¿una teoría que satisfaga la Convención T será adecuada para un lenguaje natural? Aquí hay dos cuestiones. Una de ellas es qué instrumentos deben tenerse o considerarse disponibles en el lenguaje de la teoría; la otra es cómo aplicar estos instrumentos al lenguaje del hablante. Mi supuesto de trabajo ha sido que sólo se puede recurrir a una teoría estándar de la cuantificación de primer orden. Ciertamente, he estado convencido por largo tiempo de que muchas aproximaciones alternativas a la semántica, mediante el empleo, por ejemplo, de lógica modal, de semántica del mundo posible, o de cuantificación sustitucional, no tienen cabida en una teoría que satisfaga los requerimientos de la Convención T. Ahora sé que esto fue un apresuramiento. La Convención T no es tan determinante como yo lo había pensado, y se abren más posibilidades de teorizaciones interesantes que las que yo imaginaba antes. Las bien conocidas virtudes de la teoría de la cuantificación de primer orden todavía proporcionan abundante motivación para incitarnos a descubrir cuánto podemos hacer con ella. En los tres próximos ensayos, reunidos bajo la idea común de la aplicación, intento la domesticación semántica de tres expresiones, recalcitrantes pero relacionadas entre sí: cita, discurso indirecto y operadores de modo.

El Ensayo 6 señala que ninguna teoría corriente de la cita es completamente satisfactoria, y propone una aproximación explícitamente demostrativa que haga de la cita un caso especial de la referencia demostrativa de las palabras respecto de otras palabras de su familia verbal.

El Ensayo 7, "Al decir *that*", se adentra en uno de los muchos tipos de oración utilizados para atribuir actitudes; la solución paratáctica sugerida tiene obvias afinidades con el tratamiento dado a la cita en el Ensayo 6. En el Ensayo 3 hay esbozos (que, creo, podrían haberse desa-

rrollado) acerca de cómo podría extenderse el análisis a las oraciones de creencia. Si esa estrategia se llevara a cabo, podría servir para proporcionar una semántica (aunque no una lógica) a las modalidades y a los contrafácticos, y un mayor número de oraciones acerca de las actitudes "proposicionales".

El Ensayo 8, "Modos y ejecuciones", pone el acento en la muchas veces olvidada distinción entre modos gramaticales por una parte y diversos tipos de fuerza ilocutoria por la otra. Sólo la primera es de interés para una teoría acerca del significado de las palabras. Aquí se sugiere un análisis paratáctico de los imperativos con el objeto de poner en su lugar a nuestro sentimiento natural de que los imperativos no tienen un valor de verdad, sin sobrepasar los recursos de una teoría de la verdad.

En el volumen que complementa a éste, *Essays on Actions and Events*, muestro cómo una teoría de la verdad puede aplicarse a una cantidad de casos problemáticos diversos: oraciones acerca de acciones y otros sucesos, modificación adverbial y enunciados causales singulares.

La tercera sección del presente libro está dedicada a la cuestión de si una teoría de la verdad para un hablante puede ser verificada sin dar por supuesta una gran parte de lo que pretende describir.

En "Interpretación radical", el Ensayo 9, como en el resto de los ensayos, sigo a Quine en la suposición de que aun restringiendo la atención a la conducta verbal que revela cuándo, y bajo qué condiciones, un hablante da crédito a una oración, no hay una forma directa de diferenciar los papeles de la creencia y del significado en la explicación de tal crédito. Para obtener consideraciones separadas de creencia y de significado es menester una teoría que aporte datos adicionales para contribuir a la interpretación de cada oración y de las actitudes que la acompañan. Sólo mediante el estudio del *patrón* de afirmación de oraciones podemos decidir qué es significado y qué es creencia.

El depender de la evidencia que, sin el auxilio de la teoría, no distingue entre las contribuciones de la creencia y del pensamiento a la conducta lingüística, requiere de un método para hacer efectiva la separación en un grado suficiente para la comunicación. En los presentes ensayos se describen y defienden los instrumentos para este fin. Pero todos ellos, en una u otra forma, se basan en el Principio de Caridad.

La frase y la idea básica proviene de Neil Wilson, *"Substances Without Substrata"*. Quine lo expresa así: "...aserciones que a la vista resultan pasmosamente falsas pueden descubrir con facilidad diferencias de lenguaje ocultas" (*Word and Object*, pág. 59). Quine aplica el principio fundamentalmente a la interpretación de las constantes lógicas.

Puesto que, a mi criterio, no puedo usar la noción de significado-estímulo de Quine como base para la interpretación de ciertas oraciones, aplico en todos los casos el Principio de Caridad. Así aplicado, nos aconseja preferir por lo general teorías de interpretación que minimicen el desacuerdo. Esa fue mi tendencia en los primeros ensayos, buscando acentuar la inevitabilidad del recurso a la caridad. Pero la minimización del desacuerdo, o maximización del acuerdo, es un ideal erróneo. El fin

de la interpretación no es el acuerdo sino la comprensión. Mi punto de vista siempre ha sido que la comprensión sólo puede asegurarse interpretando en una forma que lleve hacia el tipo correcto de acuerdo. El "tipo correcto", sin embargo, no es más fácil de especificar que una buena razón para sostener una creencia particular.

Las sutiles presiones sobre el Principio de Caridad comienzan a emerger en los ensayos 10 y 11. Pero aquí son todavía un asomo; en trabajos que ahora estoy preparando intento desarrollar el tema en mayor detalle.

El Ensayo 10, "La creencia y el fundamento del significado", insiste acerca de la simetría de la creencia y del pensamiento en la exploración de la conducta verbal. Pero en un aspecto importante va más allá. Desarrolla un impactante paralelo entre las teorías bayesianas de la decisión y las teorías del significado, y da razones por las cuales las dos teorías deberían considerarse mutuamente dependientes. Las insinuaciones que aparecen aquí, respecto de una futura teoría unificada del habla y la acción, fueron retomadas en mis *Carus Lectures*, y serán publicadas próximamente.

Los primeros dos ensayos sobre la interpretación radical señalan el hecho de que la comprensión de las palabras de un hablante requiere gran conocimiento de sus creencias. El Ensayo 11, "Pensamiento y habla", se ocupa de la dependencia recíproca, y concluye, en forma bastante especulativa, que sólo de una criatura con un lenguaje puede decirse con propiedad que posee un esquema maduro de actitudes proposicionales.

El Ensayo 12, "Respuesta a Foster", tal como se señaló antes, reconoce que si una teoría de la verdad pretende satisfacer los requerimientos de la interpretación, debe ser más que verdadera: sus axiomas y teoremas deben ser leyes naturales. Si un intérprete conociera una teoría así, podría emplearla para comprender a un hablante; pero sólo si supiera que los enunciados de la teoría son nómicos.

Los cuatro ensayos siguientes pueden describirse como subproductos filosóficos de la aproximación a la verdad y a la interpretación aquí sugerida.

Una teoría de la verdad puede considerarse una teoría de correspondencia en el sentido modesto del Ensayo 3, pero tal sentido no promueve la idea de que comprendamos cómo sería comparar las oraciones con aquello de que éstas hablan, puesto que la teoría no proporciona entidades con las cuales comparar las oraciones. En una línea parecida, el Ensayo 13, "De la idea misma de un esquema conceptual", explora la inteligibilidad de suponer que los diferentes lenguajes o esquemas conceptuales "dividen" o "administran" la realidad de modos sumamente diferentes. Nuestro método general de interpretación impide la posibillidad de descubrir que otros poseen un andamiaje intelectual totalmente distinto. Pero, y esto es más importante, se aduce que si rechazamos la idea de una fuente de evidencia no interpretada no queda espacio para un dualismo de esquema y contenido. Sin este dualismo no podemos dar sentido a un relativismo conceptual. Esto no significa que debamos de-

jar de lado la idea de un mundo objetivo independiente de nuestro conocimiento del mismo. El argumento contra el relativismo conceptual demuestra más bien que el lenguaje no es una cortina o filtro a través del cual debe pasar nuestro conocimiento del mundo.

Dejar de lado el dualismo de esquema y contenido equivale a abandonar un tema central del empirismo en sus principales manifestaciones históricas. Pero yo no creo que, como han sugerido amigos y críticos varias veces, mi argumento contra el empirismo me convierta, o deba convertirme, en un pragmatista, un idealista trascendental, o un realista "interno". Todas estas posiciones son formas de relativismo que encuentro tan difíciles de comprender como los empirimos que ataco.

De acuerdo con el Ensayo 13, no se puede encontrar sentido a la idea de que las fuentes conceptuales de los diferentes lenguajes difieran grandemente. El argumento que da pie a esta conclusión da pie igualmente a la conclusión de que los lineamientos generales de nuestra visión del mundo son correctos; individualmente y grupalmente podemos equivocarnos mucho, pero sólo a condición de que a grandes rasgos vayamos por buen camino. De allí que, cuando estudiamos lo que nuestro lenguaje —cualquier lenguaje— necesita en términos de ontología global, no sólo estamos pasando revista a nuestra propia imagen de las cosas: lo que consideramos que es es, en gran medida lo que es. Este es el tema del Ensayo 14, "El método de la verdad en la metafísica".

Una teoría de la verdad es puesta a prueba por teoremas que estipulan las condiciones bajo las cuales las oraciones son verdaderas; estos teoremas no dicen nada acerca de la referencia. El Ensayo 15, "Realidad sin referencia", sostiene por consiguiente que la forma en que una teoría de la verdad mapee expresiones no oracionales con objetos es indiferente en tanto que las condiciones de verdad no se vean afectadas. A qué objeto se refiere una oración particular, o a qué objeto se refiere un término, o para qué objetos es verdadero un predicado, son preguntas que no tienen respuesta.

En el Ensayo 15 coincido con Quine en sostener que la referencia es inescrutable. El Ensayo 16, "La inescrutabilidad de la referencia", advierte que no se debe considerar la inescrutabilidad como razón para tratar de relativizar en alguna medida a la referencia y a la ontología de los términos y predicados singulares. Pues, dado que nada puede revelar de qué manera un hablante ha relacionado sus palabras con los objetos, no hay nada pasible de ser relativizado; y al no verse afectada la interpetación, no hay necesidad de relativizar.

Ninguna discusión de las teorías del significado puede dejar de tomar en cuenta los límites de aplicación de dichas teorías. La esfera de acción debe ser lo suficientemente amplia como para proporcionar una visión de las formas en que el lenguaje puede servir a nuestros interminables propósitos, y lo suficientemente restringida como para ser pasible de una sistematización seria. El Ensayo 8 dio un paso necesario al distinguir entre el modo gramatical, que debe ser tenido en cuenta hasta por la más mediocre de las teorías, y la fuerza de las emisiones, que está fuera del alcance de una reglamentación comparable. El Ensayo 17,

"Qué significan las metáforas", está dedicado fundamentalmente a la tesis de que nosotros explicamos lo que las palabras hacen en la metáfora, simplemente suponiendo que ellas tienen los mismos significados que tienen en contextos no figurativos. Al postular significados metafóricos, perdemos nuestra capacidad de explicar la metáfora y eliminamos toda esperanza de una teoría responsable.

El Ensayo 18, "Comunicación y convención", delinea una nueva frontera. El desempeño de una teoría aplicada por un intérprete en un encuentro lingüístico permanece siempre como una cuestión abierta. En la práctica, el intérprete lleva adelante la conversación ajustando su teoría según el momento. Los principios de ese acomodamiento inventivo no son en sí mismos reducibles a teoría, ya que involucran nada menos que todas nuestras habilidades de construcción de teorías.

Los ensayos han sido ligeramente retocados para reducir repeticiones, eliminar pasajes innecesarios o confusos, o adecuar reflexiones originarias a modificaciones posteriores. Estas alteraciones se han limitado a lo trivial. Donde mis errores o deslices han llamado la atención he dejado todo como estaba, o señalado el cambio con una nota a pie de página.

He recibido ayuda de muchas más personas que las que aquí me es posible agradecer, pero quiero mencionar especialmente a Paul Grice, Gilbert Harman, Saul Kripke, David Lewis, Richard Rorty, Sir Peter Strawson y Bruce Vermazen. Sue Larson y Akeel Bilgrami cumplieron una tarea indispensable con las notas a pie de página, la bibliografía y el índice. Mucho más que eso, me proporcionaron consejo filosófico y apoyo moral. Sue Larson me ha enseñado mucho acerca de la filosofía del lenguaje; su influencia es especialmente fuerte en los ensayos 8 y 18. En 1970 dicté las conferencias sobre John Locke en Oxford. El contenido de aquellas conferencias surge aquí (muy modificado) en los ensayos 2, 3, 6, 7 y 13. Otra conferencia sobre la modificación adverbial aportó material hoy publicado en los ensayos 6 al 11 de *Essays on Actions and Events*.

Mi pensamiento recibió una temprana influencia de Michael Dummett, quien dictó varias conferencias sobre Frege y la filosofía del lenguaje en la Stanford University a la que yo concurría en los años cincuenta. Nuestras discusiones tomaron forma pública en 1974 cuando dictamos un seminario conjunto sobre la verdad en el All Souls College, en el que yo ejercía como profesor invitado.

A lo largo de los años, John Wallace y yo conversamos interminablemente acerca de los temas presentados en este libro. El apreció desde un principio el valor de los trabajos de Tarski sobre la verdad, y mucho de lo que yo he escrito refleja su visión y su penetrante crítica.

W. V. Quine fue mi maestro en una etapa fundamental de mi vida. No sólo me impulsó a pensar acerca del lenguaje, sino que fue el primero en hacerme ver que, en filosofía, existe el estar en lo cierto, o al menos el equivocarse, y que ello importa. Sin la inspiración de sus trabajos, de su paciente tutela, de su talento amable y de su estímulo generoso, este libro no sería peor de lo que es. Este libro no sería.

Verdad y significado

1

Teorías del significado y lenguajes aprendibles

A los filósofos les encanta hacer observaciones respecto de las propiedades que debe tener un lenguaje para poder ser, aun en principio, aprendible. El objeto de estas observaciones generalmente ha sido apoyar o socavar alguna doctrina filosófica, sea ésta epistemológica, metafísica, ontológica o ética. Pero si los argumentos son buenos deben tener implicaciones para la ciencia empírica de la formación de los conceptos, por lo menos diciendo cuáles son los límites de lo empírico.

A menudo se afirma o implica que consideraciones puramente a priori bastan para determinar rasgos de los mecanismos, o de las etapas, del aprendizaje del lenguaje; tales pretensiones son sospechosas. En la primera parte de este trabajo examino un ejemplo típico de esta posición, y hago la prueba de separar lo que podría ser aceptable de lo que no. A diferencia de algunas opiniones balbuceantes acerca de cómo aprendemos un lenguaje, propongo algo que me parece un aspecto claramente necesario de un lenguaje aprendible: debe ser posible dar una explicación constructiva del significado de las oraciones en el lenguaje. Llamo a esa explicación una teoría del significado para el lenguaje y sugiero que una teoría del significado que entre en conflicto con esta condición, haya sido presentada por un filósofo, una lingüista o un psicólogo, no puede ser una teoría de un lenguaje natural; y si ignora esta condición, deja de abordar algo esencial al concepto de un lenguaje. Sin embargo, como trato de demostrar en la segunda parte de este trabajo, una cantidad de teorías del significado usuales choca con esta condición o la ignora.

I

Primeramente aprendemos unos pocos sujetos y predicados que se aplican a objetos físicos queribles o comestibles de tamaño medio en el primer plano de los sentidos o del interés; el aprendizaje tiene lugar mediante un proceso condicionante en el que interviene la ostensión. Luego siguen predicados complejos y términos singulares para objetos no necesariamente ya observados, o permanentemente fuera de vista debido al tamaño, fecha, atenuación o inexistencia. A continuación siguen términos teóricos, quizás aprendidos mediante "postulados de significado" o a fuerza de hallarse enclavados en un discurso apropiadamente científico.

En algún punto temprano de este juego se da el gran salto del término a la oración, si bien cómo se lo da puede resultar oscuro al diluirse la transición por la existencia de oraciones de una sola palabra: "Mamá", "Fuego", "Baba", "Cabeza", etcétera.

De esta manera, en una breve caricatura, se apilan los bloques de la teoría del aprendizaje del lenguaje, reflejando, en polvorientos capítulos, la epistemiología empírica.

Esta teoría hoy ha perdido crédito en casi todos sus detalles. Por una parte, no hay razones obvias para pensar que el orden del aprendizaje se relaciona con la prioridad epistemológica. Por otra parte, algunas de sus afirmaciones parecen contradecir a la experiencia: por ejemplo, un niño aprende los términos generales "gato", "camello", "mastodonte" y "unicornio" en lo que podría ser, digamos, una manera uniforme (tal vez hojeando un libro ilustrado), si bien la relación del niño con las extensiones de estos términos es completamente diferente. En algunos casos, se puede sostener que el orden de aprendizaje del lenguaje es inverso al orden epistemológico: los datos de los sentidos pueden ser la base de nuestro conocimiento de los objetos físicos, pero se aprende a hablar de los *sense-data* —si es que se aprende— mucho después de que se consigue hablar de los objetos físicos. Por último, la epistemiología subyacente, con sus supuestos de psicología asociacionista y su simple y reduccionista teoría del significado, ya no resulta atractiva para la mayoría de los filósofos. A la luz de todo esto es asombroso que algo como la doctrina del aprendizaje del lenguaje, que comenzó como una endeble rama del empirismo temprano, florezca hoy mientras el tronco paterno se marchita. Lo que sigue es un único ejemplo de dependencia de esta doctrina obsoleta, pero este ejemplo bien podría multiplicarse en la literatura de hoy.

P. F. Strawson ha destacado la conocida opinión de Quine de que "toda la categoría de los términos singulares es teóricamente superflua".[1] Strawson concede, al menos en favor del argumento, que dentro de un lenguaje que ya contiene términos singulares, podemos parafrasear "todo cuanto hoy decimos mediante el uso de términos singulares, con formas de palabras que no contengan términos singulares" (pág. 434). Lo que Strawson niega es que de esta suposición se infiera la posibilidad teórica de que podamos hablar un lenguaje sin términos singulares, "... en el cual nunca los habríamos usado, en el cual la categoría de los términos singulares simplemente no existiera, pero en el cual seríamos capaces, no obstante, de decir todo lo que ahora somos capaces de decir mediante el uso de términos singulares" (págs. 433 y 434). Strawson luego se aboca a establecer independientemente la imposibilidad teórica de tal lenguaje.

[1] P. F. Strawson, "Singular Terms, Ontology and Identity". Referencias subsiguientes dan los números de página entre paréntesis. La cita de Quine está en la página 211 de *Methods of Logic*.

Para poder concentrarnos en el tema en discusión, permítaseme explicar que no tengo interés en responder a dos de las tesis de Strawson: éstas son, a grandes rasgos, que la *eliminabilidad* de los términos singulares no se infiere de la *parafraseabilidad* y que la eliminabilidad, tal como se la describe, es imposible. Ambas cuestiones son discutibles, desde mi punto de vista, pendientes de una aclaración de la noción de ser capaz de decir la misma cosa.

Al no estar detrás de conclusiones, entonces, mi interés se centra exclusivamente en el argumento principal que Strawson utiliza en el intento de desacreditar la tesis de eliminabilidad de Quine. Las siguientes son dos ideas esenciales de este argumento:

(1) Para que los predicados puedan ser comprendidos, algunos predicados deben aprenderse ostensivamente o por "confrontación directa";

(2) para que dicho aprendizaje tenga lugar, la situación de aprendizaje ostensivo debe estar "articulada en el lenguaje" por un elemento demostrativo que elija o identifique entidades de la clase a la cual se aplica el predicado (págs. 445 y 446).

Quine ha contestado que (1) y (2) no bastan para establecer la necesidad de términos singulares, pues los demostrativos pueden interpretarse como términos generales.[2] Esto bien puede ser verdad; mi interés, sin embargo, reside en que (1) y (2) implican que las cuestiones sustantivas acerca del aprendizaje del lenguaje pueden resolverse sobre fundamentos puramente a priori.[3]

Resumiendo su propio argumento, Strawson dice: "Algunos términos universales deben estar conectados con nuestra experiencia, para que éstos puedan ser comprendidos. Y estos términos universales deben estar relacionados con porciones o segmentos particulares de nuestra experiencia. De ahí que, para que éstos puedan aprenderse como *predicados de particulares*, deben aprenderse como predicados de particulares demostrativamente *identificados*". (pág. 446.) En bastardilla en el original. Aquí resulta quizás obvio que la noción de aprendizaje aparece vacuamente en la conclusión; volvamos entonces brevemente a (1) y (2).

Sin duda se trata de una pregunta empírica si, como resultado de ciertas experiencias, una persona pasa a tener alguna capacidad que antes no tenía; pero (1) y (2) consideran como una cuestión puramente "lógica" que cualquiera que haya adquirido una capacidad lingüística de tipo específico haya recorrido un camino prescrito. Strawson aparente-

[2] W. V. Quine, *Word and Object*, 185.

[3] Strawson cita a Quine, *Mathematical Logic*, 218, en favor de (1), y el mismo Quine puede hoy ser citado en favor de algo como (1) y (2) en *Word and Object* y en *The Roots of Reference*. Pero aparte del hecho de que Quine se niega a apoyar la conclusión de Strawson, se mantiene la importante diferencia de que Quine al menos piensa que (1) y (2) son verdaderos, mientras que Strawson quiere mostrar que son necesarios. Sólo me interesa aquí la pretensión de necesariedad, si bien dudo de que (1) ó (2) sean verdades importantes respecto de la adquisición del lenguaje.

mente iguala el aprendizaje ostentivo de un predicado con el aprendizaje por "confrontación directa". Uno puede imaginar dos motivos por los cuales este proceso se considera como algo más especial que aprender el significado de un predicado escuchando oraciones que lo asocien a términos demostrativos singulares. Uno es que el aprendizaje ostensivo puede requerir por parte del maestro una intención de que determinado objeto llame la atención del alumno. Sin embargo, parece que tal intención no es necesaria, y en efecto la mayor parte del aprendizaje de un lenguaje probablemente se debe más a la observación e imitación por parte del alumno que a cualquier propósito didáctico por parte de aquéllos observados e imitados. Una segunda diferencia es que la confrontación directa (y probablemente la ostensión, al menos tal como la interpreta Strawson) requiere la presencia de un objeto apropiado, en tanto que el uso correcto de un término demostrativo singular no la requiere. Pero no es una verdad a priori, y probablemente ni siquiera sea una verdad, que una persona no podría aprender su primer lenguaje en un entorno hábilmente falsificado. Pensar de otra manera sería, como dice Strawson en un libro posterior pero quizá muy relacionado con esto, "limitar demasiado el poder de la imaginación humana".[4]

Al defender (2) Strawson revela lo que, a mi parecer, es la confusión que subyace al argumento que he estado criticando. Según lo que se ve, no hay ninguna razón por la cual el aprendizaje ostensivo de predicados debe estar "articulado en el lenguaje" de una forma en lugar de otra; no hay ninguna razón por la cual la referencia a particulares (suponiendo que esto sea necesario) deba darse por medio de demostrativos. Tampoco Strawson proporciona razón alguna; en cambio, da fundamentos a la pretensión de que "ningún simbolismo puede ser interpretado como un lenguaje en el cual se hace referencia a los particulares, a menos que contenga instrumentos para hacer referencias demostrativas... a particulares, esto es, a menos que contenga, por ejemplo, términos singulares para referirse a particulares...".[5] En el mismo trabajo, más adelante Strawson ataca "... la presunción acrítica de que sólo una parte de la estructura del lenguaje ordinario podría existir y funcionar aislada del todo del cual es una parte, tal como funciona cuando está incorporada a ese todo... Hasta este punto, al menos, el lenguaje es orgánico" (451 y 452). Aquí emerge la idea, desprovista de cualquier conexión importante con el aprendizaje del lenguaje, de que la tesis de eliminabilidad de Quine es falsa pues efectuaríamos un cambio concepual de envergadura —alterar los significados de todas las oraciones retenidas— si aisláramos a éstas de sus relaciones presentes con oraciones que contienen términos singulares (o demostrativos, o nombres propios, etcétera). Sugiero que, partiendo de argumentos en favor de la interdependencia conceptual de diversas formas idiomáticas bási-

[4] P. F. Strawson, *Individuals*, 200.
[5] P. F. Strawson, "Singular Terms, Ontology and Identity", 447.

cas, Strawson ha extraído ilegítimamente conclusiones respecto del mecanismo y secuencia de adquisición del lenguaje.

Con respecto a la interdependencia conceptual, Quine parece concordar, al menos hasta cierto punto, y escribe que "el término general y el singular demostrativo son, junto con la identidad, instrumentos interdependientes que el niño de nuestra cultura debe domesticar en un solo revoltillo alocado".[6] Si estas pretensiones de interdependencia son defendibles, sigue siendo cierto que no podríamos aprender un lenguaje en el cual los predicados significaran lo que significan los predicados en nuestro lenguaje y en el cual no hubiera términos singulares demostrativos. No podríamos aprender tal lenguaje pues, de acuerdo con los supuestos, no podría haber un lenguaje así.[7] La lección totalmente negativa para las teorías de aprendizaje del lenguaje, aunque no carente de importancia: en la medida en que tomemos seriamente el carácter "orgánico" del lenguaje, no podemos describir con exactitud los primeros pasos hacia su conquista como un aprender parte del lenguaje; es más bien una cuestión de aprendizaje en parte.[8]

II

No resulta apropiado esperar que consideraciones lógicas dicten la vía o el mecanismo de adquisición del lenguaje, pero estamos autorizados a considerar anticipándonos al estudio empírico lo que supondremos por conocer un lenguaje, cómo describiremos la habilidad o capacidad de una persona que ha aprendido a hablar un lenguaje. Una condición natural a imponer es que debemos ser capaces de definir un predicado de expresiones, basado únicamente en las propiedades formales de éstas, que revele la clase de las expresiones significativas (oraciones), en base al supuesto de que diversas variantes psicológicas se mantienen constantes. Este predicado proporciona la gramática del lenguaje. Otra condición, aun más interesante, es que debemos ser capaces de especificar, de un modo que dependa efectiva y únicamente de consideraciones formales, lo que cada oración significa. Con los accesorios psicológicos adecuados, nuestra teoría debería equipararnos para determinar en una oración arbitraria lo que el hablante del lenguaje significa con esa oración (o lo que cree que ella significa). Con la guía de una teoría adecuada vemos cómo las acciones y disposiciones de los hablantes inducen una estructura semántica en la oraciones del lenguaje. Si bien indudablemente relativizado a tiempos, lugares y circunstancias, el tipo de estructura requerido parece idéntico o muy similar al tipo que surge de

[6] W. V. Quine, *Word and Object*, 102.

[7] En *Individuals*, Strawson ataca nuevamente lo que él considera la tesis de eliminabilidad de Quine, pero no utiliza el argumento del aprendizaje que yo estoy criticando.

[8] Este tema se retoma en el Ensayo 11. En el Ensayo 8 de *Essays on Action and Events* discuto la cuestión de si los objetos son conceptualmente previos a los hechos.

una definición de la verdad según la línea iniciada por Tarski, pues esta definición proporciona un efectivo método para determinar el significado de cada oración (por ejemplo, enumera las condiciones bajo las cuales es verdadera).[9] No intento sostener aquí la necesidad de ser capaces de extraer una definición de la verdad a partir de una teoría adecuada (si bien se requiere algo así), pero una teoría cumple la condición que yo tengo en mente si podemos extraer de ella una definición de la verdad; no se requiere en particular una noción de significado más fuerte que ésta.[10]

Estas cuestiones parecen estar relacionadas con la posibilidad de aprender un lenguaje en una manera informal que ahora describo. Cuando podemos considerar el significado de cada oración como una función de un número finito de rasgos de la oración, no sólo tenemos una percepción de lo que se puede aprender; también comprendemos cómo una aptitud infinita puede abarcarse mediante logros finitos. Y bien, supongamos que un lenguaje careciera de esta característica; entonces, no importa cuántas oraciones aprenda a generar y comprender el candidato a hablante, siempre quedarán otras cuyos significados no respondan a las reglas ya conocidas. Es natural concluir que un lenguaje así es inaprendible. Este argumento depende, desde luego, de un número de supuestos empíricos: por ejemplo, que no adquirimos repentinamente en algún momento una capacidad para intuir sin regla alguna los significados de oraciones; que cada nuevo ítem de vocabulario, o cada nueva regla gramática, insume un tiempo finito para su aprendizaje; que el hombre es mortal.

Llamemos *primitivo semántico* a una expresión para la cual las reglas que determinan el significado de las oraciones en que ella no aparece no bastan para determinar el significado de las oraciones en que ella sí aparece. Entonces podemos enunciar la condición en discusión diciendo: un lenguaje aprendible tiene un número finito de primitivos semánticos. Por rústico que sea este enunciado de la condición, creo que es lo suficientemente claro como para sostener la idea de que una cantidad de teorías del significado recientes no son, ni siquiera en principio, aplicables a los lenguajes naturales, pues los lenguajes para los cuales tiene aplicación no son aprendibles en el sentido descrito. Paso ahora a dar ejemplos.

Primer ejemplo. El signo de una cita: las comillas. Las comillas deberían intrigarnos más de lo que lo hacen. Comprendemos muy bien las comillas, al menos en el sentido de que siempre comprendemos la referencia a una cita. Puesto que el número de citas es infinito, aparentemente nuestro conocimiento atesora una regla. La complicación aparece

[9] A. Tarski, "The Concept of Truth in Formalized Languages".

[10] Muchos autores, incluso Tarski, han pensado que era imposible dar una definición de la verdad hasta para las oraciones indicativas de un lenguaje natural. Si esto es así, debemos encontrar otra manera de mostrar cómo los significados de las oraciones dependen de sus estructuras. Véanse los Ensayos 2, 4 y 9.

cuando tratamos de expresar esta regla como un fragmento de una teoría del conocimiento.

Es fácil dar con aproximaciones informales de lo que entendemos por ello. Quine dice: "Comúnmente se indica el nombre de un nombre o de otra expresión colocando la expresión nombrada entre comillas simples... del todo, llamado *cita*, denota su interior";[11] Tarski dice esencialmente lo mismo.[12] Es obvio que estas fórmulas ni siquiera proporcionen el núcleo de una teoría en el sentido querido, como ambos autores se esfuerzan en señalar. Quine destaca que las citas tienen "cierta característica anómala" que "requiere una especial cautela"; Church llama "engañoso" a este recurso. Lo engañoso es, quizás, que nos vemos tentados a considerar a cada par de comillas como una expresión funcional pues cuando encierra a una expresión el resultado denota dicha expresión. Pero para despojar esta idea, debemos tratar a las expresiones entre comillas como términos o variables singulares. Tarski señala que, en casos favorables, se llega a paradojas; en casos desfavorables, la expresión comprendida carece de significado (págs. 159-62). Debemos entonces dejar de lado la idea de que las citas son "expresiones sintácticamente compuestas, de las cuales son parte tanto las comillas como las expresiones que ellas encierran". La única alternativa que nos ofrece Tarski es la siguiente:

> Los nombres de una cita deben tratarse como palabras simples de un lenguaje y por lo tanto como expresiones sintácticamente simples. Los constituyentes simples de estos nombres —las comillas y las expresiones comprendidas entre ellas— cumplen la misma función que las letras y los complejos de letras sucesivas en las palabras simples. De allí que no puedan poseer un significado independiente. Cada nombre de una cita es así un nombre individual constante de una expresión definida... y en efecto un nombre del nombre naturaleza como el nombre propio de un hombre (pág. 159).

Aparentemente en la misma línea de pensamiento, Quine escribe que una expresión entre comillas "se encuentra allí como un mero fragmento de un nombre más largo que contiene, aparte de este fragmento, las dos comillas", y compara la presencia de una expresión entre comillas con la presencia de "cat" en "cattle"[13] y de "can" en "canary".[14]

La función de las letras en las palabras, como la función de "cat" en "cattle", es puramente accidental en este sentido: podríamos sustituir "cattle" por un nuevo grupo tipográfico en cualquier punto del lenguaje y nada se vería modificado en la estructura semántica del lenguaje. No sólo "cat" no tiene un "significado separado" en "cattle"; el hecho de que las mismas letras se presenten juntas y en el mismo orden en

[11] W. V. Quine, *Mathematical Logic*, nº 4.
[12] A. Tarski, "The Concept of Truth in Formalized Languages", 156.
[13] W. V. Quine, *From a Logical Point of View*, 140.
[14] W. V. Quine, *Word and Object*, 144.

cualquier parte es irrelevante en cuestiones de significado. Si una afirmación análoga es válida para las citas, luego no hay justificación en teoría para la clasificación (es sólo un accidente que las citas compartan una característica común en su deletreo), y el hecho de que una cita nombre a "su interior" no tiene importancia. Finalmente, cada cita es un primitivo semántico, y, puesto que hay infinitas citas diferentes, un lenguaje que contiene citas es inaprendible.

Esta conclusión choca contra nuestra intuición. No hay ningún problema en delinear una regla general para identificar citas en base a la forma (toda expresión encerrada entre comillas), ni lo hay en dar una regla informal para producir una cita deseada (encerrar la expresión que uno quiere mencionar entre comillas). Puesto que estas reglas implican que las citas tienen estructura significativa, es difícil negar que deba haber una teoría semántica que la explote. Tampoco es del todo manifiesto que Tarski o Quine quieran negar esa posibilidad. Tarski considera sólo dos análisis de citas, pero no descarta explícitamente otras, como no respalda abiertamente la alternativa que no rechaza. Y ciertamente parece insinuar que las citas tienen una estructura significativa cuando dice: "Es claro que podemos correlacionar un nombre estructural-descriptivo con cada nombre-comilla, uno que no está sujeto a comillas y posee la misma extensión (por ejemplo, denota la misma expresión) y viceversa" (pág. 157). Es difícil entrever cómo podría establecerse la correlación si reemplazáramos cada cita por algún otro símbolo arbitrario, como podríamos hacer si los nombres-comilla fueran como los nombres propios de los hombres.

Quine lleva las cosas algo más allá cuando afirma que si bien las citas son "lógicamente inestructuradas" (pág. 190), y desde luego las expresiones se presentan de manera no referencial dentro de las comillas, aún esta última característica puede "desvanecerse por un sencillo cambio de notación" (pág. 144) lo que nos deja con los recursos lógicamente estructurados del deletreo y la concentración. La fórmula para el "sencillo cambio" aparentemente puede darla una definición (págs. 189, 190). Si esta sugerencia puede llevarse adelante, luego el aspecto más recalcitrante de la cita da paso a la teoría, pues las condiciones de verdad para oraciones que contienen citas pueden igualarse a las condiciones de verdad para oraciones obtenidas de ellas sustituyendo las citas por sus equivalentes definicionales en la fórmula ideomática del deletreo. En una teoría así ya no hay un número infinito de primitivos semánticos, a pesar del hecho de que no pueda demostrarse que las citas contengan partes con roles semánticos independientes. Si aceptamos una teoría de este tipo, nos vemos forzados a admitir una especie de estructura que puede no merecer ser llamada "lógica", pero que sin duda está directa e indisolublemente ligada a la estructura lógica, un tipo de estructura que falta en los nombres propios usuales.[15]

[15] Peter Geach ha insistido, en *Mental Acts* y demás, que una cita "debe considerarse *describiendo* la expresión en términos de sus partes" (*Mental Acts*, 83). Pero no explica có-

Segundo ejemplo. Scheffler y el discurso indirecto. Israel Scheffler ha propuesto lo que él llama una aproximación inscripcional al discurso indirecto.[16] Carnap en una época analizó las oraciones del discurso indirecto considerando una relación (elaborada en términos de una noción de isomorfismo intensional) entre un hablante y una oración.[17] Church objetó que el análisis de Carnap conllevaba una referencia implícita a un lenguaje que faltaba en las oraciones a ser analizadas, agregando que creía que cualquier análisis correcto interpretaría que la cláusula-*that** hace referencia a una proposición.[18] Scheffler se abocó a mostrar que un análisis correcto podría arreglárselas en menos, demostrando lo que podía hacerse con una ontología de inscripciones (y de emisiones).

Scheffler piensa que analizamos "Tonkin dijo *that* la nieve es blanca" como "Tonkin habló una emisión *that*-la-nieve-es-blanca". Dado que una emisión o inscripción pertenece eternamente al lenguaje de su productor efímero, el reproche que Church hace a Carnap carece aquí de fuerza. La parte esencial de la historia, desde el punto de vista del interés actual, es que la expresión "*that*-la-nieve-es-blanca" sea tratada como un predicado unitario (de emisiones o inscripciones). Para Scheffler el *"that"* es un operador que se aplica a una oración para formar un término general compuesto. En este uso, "compuesto" no puede significar "lógicamente complejo"; al menos, mientras no dispongamos de más teoría. Como en el caso de las citas, la sintaxis es lo suficientemente clara (como por lo general lo es en el discurso indirecto, con alguna modificación en verbos o pronombres en base a principios más fáciles de dominar que de describir). Pero no hay ninguna alusión respecto de cómo el significado de estos predicados depende de su estructura. Si una teoría falla, debemos ver en cada nuevo predicado un primitivo semántico. Sin embargo, dada su sintaxis (póngase cualquier oración detrás de *"that"* y sazónesela con guiones), es obvio que hay infinitos predicados así, por lo cual los lenguajes sin más estructura que la que Scheffler admite son, a mi entender, inaprendibles. Incluso la idea de que *"that"* es un operador de formación de predicados debe reconocerse como una observación puramente sintáctica que no tiene eco en la teoría del significado.

Podría contestarse que permanece abierta la posibilidad de generar una teoría que revele más estructura. Es cierto. Pero si Scheffler quiere el beneficio de esta posibilidad, debe pagar por él renunciando a la pretensión de haber mostrado *"la forma lógica y el carácter ontológico"* de las oraciones del discurso indirecto;[19] más teoría puede significar

---

mo se puede asignar a una cita la estructura de una descripción. Para una discusión más amplia de la cita véase el Ensayo 6.

* El autor utiliza el término inglés *that* en sus dos sentidos, que en castellano se traducen por "que" y "eso", en este capítulo y en referencias más adelante. [T.]

16 I. Scheffler, "An Inscriptional Approach to Indirect Discourse" y *The Anatomy of Inquiry.*

17 R. Carnap, Meaning and Necessity.

18 A. Church, "On Carnap's Analysis of Statements of Assertion and Belief".

19 I. Scheffler, *The Anatomy of Inquiry*, 101. (La bastardilla es suya.)

más ontología. Si la forma lógica nos habla de ontología, entonces el ardid de las citas de Quine no funcionará, pues la forma lógica de una cita es la de un término singular sin partes, y su ontología manifiesta es tan solo la expresión que nombra. Pero la definición que aprovecha su estructura presenta más entidades (aquellas nombradas en el deletreo). Es sólo un accidente que en esta aplicación el método no aporte entidades que aún no están nombradas por alguna cita.[20]

Tercer ejemplo. Quine y las oraciones de creencia. Siempre podemos cambiar problemas de ontología surgidos a raíz de expresiones y posiciones putativamente referenciales, por problemas de articulación lógica. Si una expresión irrita por su supuesta referencia, no es necesario extirparla. Basta con declararla una parte sin sentido de una expresión significativa. Este tratamiento no será matar antes de curar; después de que todos los términos y posiciones singulares pasibles de cuantificación hayan sido fijados a sus contextos, quedará la estructura lógica creada por el tratamiento de las oraciones como unidades inanalizables y por los conectivos oracionales puros. Pero la semántica sin ontología no es muy interesante, y un lenguaje como el nuestro para el cual no pudiera hacerse algo mejor sería un paradigma de inaprendibilidad.

Quine, desde luego, no ha llegado tan lejos, pero por un aisado momento parece haberse acercado a ello. Próximo a la conclusión de una prolongada, brillante y desalentadora búsqueda de una teoría satisfactoria de las oraciones de creencia (y de sus parientes cercanos: oraciones del discurso indirecto, oraciones acerca de dudas, asombros, miedos, deseos, etcétera) Quine afirma que una vez que renunciamos a tratar de cuantificar respecto de las cosas creídas:

> ... no hay ninguna necesidad de reconocer a "cree" y a verbos similares como términos emparentados; no hay necesidad de sostener su uso predicativo como en "*w* cree *x*" (comparado con "*w* cree que *p*"); no hay necesidad, por lo tanto, de ver a "que *p*" como un término. De allí, una alternativa final que yo encuentro tan atractiva como cualquier otra es simplemente hacer caso omiso a los objetos de las actitudes proposicionales... Esto significa considerar a "Tom cree (Cicerón denunció a Catilina)"... como de la forma "F*a*" donde *a* =Tom y "F" es compleja. El verbo "cree" cesa aquí de ser un término y se vuelve parte de un operador "cree que"... el cual, aplicado a una oración, produce un término general absoluto compuesto del cual la oración se considera un constituyente inmediato.[21]

En un aspecto, esto va un paso más allá que Scheffler, pues hasta el verbo principal ("cree" en este caso) se hace inaccesible al análisis lógico. La referencia a constituyentes y operadores debe ser tomada, por supuesto, como puramente sintáctica, sin apoyo en la teoría semántica.

[20] Para mayor discusión del discurso indirecto, véase el Ensayo 7.
[21] W. V. Quine, *Word and Object*, 215, 216.

Si hay algún elemento común a las oraciones de creencia en tanto clase más allá de la sintaxis, el informe de Quine no dice cuál es. Y, desde luego, un lenguaje para el cual no puede darse más teoría es, a mi entender, inaprendible. Quine, sin embargo, no descarta la esperanza de que surja más teoría, quizás algo parecido a lo que ofrece respecto de las citas. Pero en el caso de las oraciones de creencia no es fácil imaginar la teoría que podría dar por fruto la estructura requerida sin añadir ontología.

Cuarto ejemplo. Church y la Lógica del Sentido y la Denotación. En los dos últimos ejemplos, la pérdida de un mínimo deseable de articulacion de significado pareciera ser el resultado de un celo excesivo en la atención de los problemas de ontología. Pero sería un error inferir que resolveríamos todos los problemas siendo pródigos con las entidades intensionales.

Frege nos pide que supongamos que ciertos verbos, como "cree" (o "cree que"), cumplen una doble función.[22] Primero, crean un contexto en el cual las palabras que siguen pasan a referir su sentido o significado usual. Segundo (suponiendo que el verbo es el verbo principal de la oración), cumplen un tipo de función normal al mapear personas y proposiciones con valores de verdad. Esta doctrina parece bastante oscura, particularmente con respecto al primer punto, y Frege parece haberlo pensado. Su punto de vista fue, tal vez, que esto es lo mejor que podemos hacer por los lenguajes naturales, pero que en un lenguaje lógicamente más transparente se usaría palabra diferentes para referirse al sentido y a la denotación, aliviando así a los verbos sobrecargados de su primera, y más oscura, tarea. Pero ahora estamos sólo a unos pasos de un vocabulario primitivo infinito. Tras las primeras apariciones de verbos como "cree" introducimos nuevas expresiones para los sentidos. Luego, notamos que no hay una teoría que interprete estas nuevas expresiones como lógicamente estructuradas. Otra vez hace falta un nuevo vocabulario, por la mismas razones, cada vez que iteramos "cree"; y no hay un límite para el número de iteraciones posibles.

No debe pensarse que habría menos problemas con la sugerencia original de Frege. Aun si supusiéramos haber dado un sentido correcto a la idea de que ciertas palabras crean un contexto en el cual otras palabras toman nuevos significados (una idea que sencillamente las hace aparecer como functores), restaría la tarea de reducir a teoría la determinación de dichos significados: un número infinito, al menos para algunas palabras. El problema no es cómo las expresiones individuales que conforman una oración gobernada por "cree", dados los significados que tienen en un contexto tal, se combinan para denotar una proposición; el problema es más bien enunciar la regla que dé a cada una el significado que efectivamente tiene.

[22] Véase G. Frege, "On Sense and Reference".

Volviendo a nuestras especulaciones en lo concerniente a la forma que Frebe habría dado a un lenguaje notacionalmente superior a lenguaje ordinario, pero equivalente a éste en su capacidad de tratar con la oraciones de creencia y demás. Digamos que las características que más arriba atribuí a este lenguaje se encuentran en un lenguaje propuesto por Alonzo Church.[23] En la notación de Church, el hecho de que las nuevas expresiones que entran en juego a medida que ascendemos en la escala semántica no sean lógicamente complejas se ve oscurecido superficialmente a raíz de que están compuestas sintácticamente por las expresiones del nivel inmediatamente inferior, más un cambio de subíndice. Todo está ideado de tal manera que "si, en una fórmula bien formada sin variables libres, todos los subíndices en todos los símbolos tipo que aparecen se incrementan en 1, la fórmula bien formada resultante denota el sentido de la primera" (pág. 17). Pero, desde luego, esta regla no puede explotarse como parte de una teoría del significado para el lenguaje; no puede tomarse la expresión y el subíndice como si tuviesen significados independientes. Respecto de una expresión en un nivel dado, las expresiones del nivel superior son primitivos semánticos, tal como Church indica claramente (pág. 8). No sugiero error alguno por parte de Church; él nunca insinúa que la cuestión sea distinta de lo que yo digo. Pero sí me permito decir que el lenguaje de Sentido y Denotación de Church es, incluso por principio, inaprendible.

[23] A. Church, "A Formulation of the Logic of Sense and Denotation".

2

Verdad y significado

La mayor parte de los filósofos del lenguaje, y recientemente algunos lingüistas, han aceptado que una teoría del significado satisfactorio debe dar cuenta de cómo los significados de las oraciones dependen del significado de las palabras. Se argumenta que, a menos que pudiera rendirse cuenta de ello para un lenguaje particular, no habría forma de explicar el hecho de que podemos aprender el lenguaje: no se podría explicar el hecho de que, a través del dominio de un vocabulario finito y de un conjunto de reglas enunciadas de manera finita, estamos preparados para producir y comprender cualquier oración de la potencial infinitud de ellas. No discuto estas vagas pretensiones, en las que veo algo más que una semilla de verdad.[1] En cambio, quiero cuestionar qué es para una teoría dar una explicación del tipo esbozado.

Una propuesta es comenzar por asignar alguna entidad como significado de cada palabra (u otra característica sintáctica significativa) de la oración; de esta manera podemos asignar Teeteto a "Teeteto" y la propiedad de volar a "vuela" en la oración "Teeteto vuela". Luego surge el problema de cómo a partir de estos significados se genera el significado de la oración. Si consideramos que la concatenación es una parte importante de la sintaxis, podemos asignarle a ella la relación de participación o instanciación; sin embargo, es obvio que tenemos aquí el comienzo de un regreso infinito. Frege trató de evitar el regreso diciendo que las entidades correspondientes a predicados (por ejemplo) están "insaturadas" o "incompletas" a diferencia de las entidades que corresponden a nombres, pero esta doctrina parece más bien nombrar una dificultad, y no resolverla.

La cuestión emergerá si por un momento pensamos en los términos singulares complejos, para los cuales es de aplicación la teoría de Frege tanto como para las oraciones. Consideremos la expresión "el padre de Annette"; ¿de qué manera el sentido del todo depende del sentido de las partes? La respuesta parecería ser que el significado de "el padre de" es tal que cuando esta expresión se prefija a un término singular el resultado refiere al padre de la persona a quien el término singular refiere. ¿Qué papel cumple, en esta interpretación, la entidad instaurada

[1] Véase el Ensayo 1.

o incompleta que está en lugar de "el padre de"? Todo lo que se nos ocurre decir es que esta entidad "ofrece" o "da" al padre de *x* como valor cuando el argumento es *x*, o quizá que esta entidad relaciona a personas con sus padres. Puede no resultar claro si la entidad que supuestamente está en lugar de "el padre de" ejecuta alguna función explicativa genuina en tanto nos limitemos a expresiones individuales; así, pensemos entonces en la clase infinita de expresiones que se forman al escribir "el padre de" cero o más veces frente a "Annette". Es fácil proporcionar una teoría que diga, para uno cualquiera de estos términos singulares, a qué refiere dicho término: si el término es "Annette" refiere a Annette, mientras que si el término es complejo y consiste en "el padre de" prefijado a un término singular *t*, en ese caso refiere al padre de la persona a quien *t* refiere. Es obvio que al formular esta teoría no se menciona, no hace falta mencionar, a ninguna entidad correspondiente a "el padre de".

No sería apropiado denunciar que esta pequeña teoría *usa* las palabras "el padre de" al dar la referencia de expresiones que contienen estas palabra. Pues la tarea consistía en proporcionar el significado de todas las expresiones de un determinado conjunto infinito en base al significado de las partes; no entraba en los cálculos proporcionar también los significados de las partes atómicas. Por otra parte, es hoy evidente que una teoría satisfactoria de los significados de expresiones complejas puede no requerir entidades como significados de todas las partes. Nos corresponde entonces replantear nuestros requisitos para una teoría del significado satisfactoria de modo de no insinuar que las palabras individuales deben tener significados, en algún sentido que trascienda al hecho de que aquéllas tienen un efecto sistemático sobre los significados de las oraciones en que aparecen. En realidad, para el caso que tenemos entre manos podemos ir aun más lejos al establecer el criterio de éxito: lo que queríamos, y lo que obtuvimos, es una teoría que implica a todas las oraciones de la forma "*t* refiere a *x*" donde "*t*" es reemplazada por una descripción estructural[2] de un término singular, y "*x*" es reemplazada por ese término mismo. Más aun, nuestra teoría logra esto sin apelar a ningún concepto semántico más allá del básico "refiere a". Finalmente, la teoría sugiere claramente un procedimiento efectivo para determinar, para todo término singular de este universo, a qué refiere dicho término.

Una teoría con semejantes méritos evidentes merece una aplicación más amplia. El recurso propuesto por Frege a este efecto es de una simplicidad brillante: tomar a los predicados como un caso especial de expresiones funcionales, y a las oraciones como un caso especial de términos singulares complejos. Sin embargo, si queremos continuar en nuestro presente (implícito) camino de identificación del significado de

[2] Una "descripción estructural" de una expresión describe a la expresión como una concatenación de elementos extraídos de una lista finita y fija (por ejemplo, de palabras o letras).

un término singular con su referencia, asoma una dificultad. La dificultad surge tras hacer dos suposiciones razonables: que términos singulares lógicamente equivalentes tienen la misma referencia, y que un término singular no cambia por otro con la misma referencia. Pero supongamos ahora que "*R*" y "*S*" abrevian a dos oraciones cualesquiera iguales en valor de verdad. Luego, las cuatro oraciones siguientes tienen la misma referencia:

(1) R
(2) $x\ (x = x.R) = x\ (x = x)$
(3) $x\ (x = x.S) = x\ (x = x)$
(4) S

Pues (1) y (2) son lógicamente equivalentes, como los son (3) y (4), mientras que (3) difiere de (2) solamente porque contiene al término sigular "$x\ (x = x.S)$" donde (2) contiene a "$x\ (x = x.R)$" y ambas refieren a la misma cosa si S y R son iguales en valor de verdad. De aquí que dos oraciones cualesquiera tengan la misma referencia si tienen el mismo valor de verdad.[3] Y si el significado de una oración es lo que ella refiere, todas las oraciones iguales en valor de verdad deben ser sinónimas: un resultado intolerable.

Aparentemente debemos abandonar este camino en nuestra búsqueda de una teoría del significado. Este es el momento adecuado para buscar auxilio en la distinción entre significado y referencia. El problema, se nos informa, es que las cuestiones de referencia las solucionan en general hechos extralingüísticos, no así las cuestiones de significado, y los hechos pueden fusionar a las referencias de expresiones que no son sinónimos. Si queremos una teoría que dé el significado (diferenciado de la referencia) de cada oración, debemos comenzar por el significado (diferenciado de la referencia) de las partes.

Hasta aquí hemos estado siguiendo los pasos de Frege; gracias a él, la senda es bien conocida e incluso bien recorrida. Pero ahora, quisiera sugerir, hemos llegado a un *impasse:* el paso de la referencia al significado no nos lleva a ninguna explicación útil del modo en que los significados de las oraciones dependen de los significados de las palabras (u otros rasgos estructurales) que las componen. Preguntemos, por ejemplo, cuál es el significado de "Teeteto vuela". Una respuesta fregeana sería algo así: dado el significado de "Teeteto" como argumento, el significado de "vuela" proporciona el significado de "Teeteto vuela" como valor. La vacuidad de esta respuesta es obvia. Queríamos saber cuál es el significado de "Teeteto vuela"; no es ningún progreso que se nos diga qué es el significado "Teeteto vuela". Esto lo sabíamos antes de tener alguna teoría a la vista. En la explicación espuria que acabamos de ver, es

[3] El argumento deriva de Frege.Véase A. Church, *Introduction to Mathematical Logic,* 24-5. Quizá valga la pena mencionar que el argumento no depende de ninguna identificación particular de las entidades a las cuales se supone que refieren las oraciones.

ocioso hablar de la estructura de la oración y de los significados de las palabras pues ello no cumplía ningún papel en la producción de la descripción dada del significado de la oración.

El contraste entre una explicación real y una explicación pretendida será aun más notorio si buscamos una teoría, análoga a la teoría mínima de la referencia de términos singulares recién delineada, pero que a diferencia de ésta trate con significados en lugar de tratar con referencias. Lo que la analogía requiere es una teoría que tenga por consecuencias a todas las oraciones de la forma "*s* significa *m*" donde "*s*" sea reemplazada por una descripción estructural de la oración y "*m*" sea reemplazada por una descripción estructural de la oración y "*m*" sea remplazada por un término singular que refiera al significado de dicha oración; más aun, una teoría que provea un método efectivo para llegar al significado de una oración arbitraria estructuralmente descrita. Evidentemente, para que estos criterios puedan cumplirse resulta esencial para referir los significados una vía más articulada que cualquiera de las que hasta ahora hemos visto.[4] Los significados como entidades, por el concepto afín de sinonimia, nos permiten formular la siguiente regla que relaciona a las oraciones con sus partes: son sinónimas las oraciones cuyas partes correspondientes son sinónimas ("correspondientes", desde luego, requiere aquí de una aclaración). Y en teorías como la de Frege, los significados como entidades pueden, en ocasiones, cumplir la función de referencias, perdiendo así su estatus de entidades diferenciadas de las referencias. Paradójicamente, una cosa que los significados no parecen hacer es allanar el camino de una teoría del significado; al menos en la medida en que exigimos a una teoría tal que dé en forma no trivial el significado de cada oración del lenguaje. Lo que yo objeto a los significados en la teoría del significado no es su condición de abstractos ni la oscuridad de sus condiciones de identidad, sino que carecen de una utilidad demostrada.

Este es el momento de introducir una nueva esperanza. Supongamos que tenemos para nuestro lenguaje una teoría satisfactoria de la sintaxis, que consiste en un método efectivo para determinar, para una expresión arbitraria, si es o no independientemente significativa (por ejemplo una oración), y pensamos como es usual que esto implica considerar que cada oración está compuesta, de una manera admisible, por elementos extraídos de una reserva finita y fija de elementos sintácticos atómicos (esto es palabras). La esperanza es que la sintaxis así concebida dará la semántica cuando agreguemos un diccionario que dé el significado de cada átomo sintáctico. La esperanza se desvanecerá, sin em-

[4] Puede pensarse que Church, en "A Formulation of the Logic of Sense and Denotation", ha presentado una teoría del significado que hace uso esencial de los significados como entidades. Pero éste no es el caso: se interpreta que las lógicas de sentido y denotación de Church tratan de los significados, pero no mencionan a las expresiones y por lo tanto no pueden ser teorías del significado en el sentido ahora en discusión.

bargo, si es que la semántica va a incluir una teoría del significado en el sentido que nosotros le damos a ésta, pues el conocimiento de las características estructurales que hacen significativa a la oración, más el conocimiento de los significados de las partes últimas, no dan por resultado el conocimiento de lo que significa una oración. Esto lo ilustran claramente las oraciones de creencia. Su sintaxis es relativamente sencilla. Pero el agregado de un diccionario no modifica el problema semántico normal, cual es que ni siquiera podemos dar cuenta de las condiciones de verdad de tales oraciones partiendo de nuestro conocimiento de los significados de las palabras que contienen. La situación no mejora mucho aunque podamos refinar el diccionario para indicar qué significado o significados asume una expresión ambigua en cada uno de sus contextos posibles; el problema de las oraciones de creencia persiste aun después de resolver las ambigüedades. El hecho de que la sintaxis recursiva sumada a un diccionario no sea necesariamente semántica recursiva se ha visto oscurecido en algunos trabajos de lingüística recientes a raíz de la intrusión de criterios semánticos dentro de la discusión de teorías sustancialmete sintácticas. Si los criterios semánticos fueran claros, la cuestión se reduciría a una inofensiva diferencia sobre terminología; pero no lo son. Si bien hay acuerdo en que la tarea principal de la semántica es dar la interpretación semántica (el significado) de cada oración del lenguaje, en ninguna parte de la literatura lingüística se podrá encontrar, al menos por cuanto yo sé, una explicación llana del modo en que una teoría realiza esta tarea, o del modo de determinar cuándo ella ha sido completada. El contraste con la sintaxis es sorprendente. La tarea principal de una modesta sintaxis es caracterizar la significatividad (u oracionalidad). Podemos tener tanta confianza en la corrección de esta caracterización como la tenemos en la representatividad del ejemplo que estamos analizando y en nuestra capacidad de decir cuándo las expresiones particulares son significativas (oraciones). ¿Qué tarea y comprobación tan claras y análogas existen para la semántica?[5]

Decidimos algo más atrás no suponer que las partes de las oraciones tienen significados, excepto en el sentido ontológicamente neutro de efectuar una contribución sistemática al significado de las oraciones en las que aparecen. Dado que postular significados no ha sido útil, volvamos a aquella idea. Una de las direcciones hacia las que apunta es una cierta visión holística del significado. Si el significado de las oraciones

[5] Para una exposición reciente acerca del papel de la semántica en la lingüística, véase Noam Chomsky, "Topics in the Theory of Generative Grammar". En este artículo, Chomsky (1) insiste en la importancia crucial de la semántica en la teoría lingüística, (2) sostiene la superioridad de las gramáticas tranformacionales sobre las gramáticas de estructura de frase sobre todo en base a que, si bien las gramáticas de estructura de frase pueden ser adecuadas para definir la oracionalidad para (al menos) algunos lenguajes naturales, resultan inadecuadas como base de la semántica, y (3) hace referencia repetidamente al "estado más bien primitivo" de los conceptos de semántica y señala que la noción de interpretación semántica "aún se resiste a cualquier análisis profundo".

depende de su estructura, y entendemos el significado de cada ítem de la estructura sólo como una abstracción de la totalidad de oraciones en las que éste aparece, entonces podemos dar el significado de cualquier oración (o palabra) solamente dando el significado de todas las oraciones (y palabras) del lenguaje. Frege dijo que una palabra tiene significado sólo en el contexto de una oración; siguiendo esta idea debería haber agregado que sólo en el contexto del lenguaje una oración (y por lo tanto una palabra) tiene significado.

Este grado de holismo ya estaba implícito en la sugerencia de que una teoría del significado adecuada debe implicar a *todas* las oraciones de la forma "*s* significa *m*". Pero, ahora que no hemos encontrado en los significados de las oraciones mayor ayuda que la que nos daban los significados de las palabras, la pregunta es si podremos deshacernos de los problemáticos términos singulares que supuestamente reemplazan a "*m*" y refieren a los significados. En un sentido, nada podría ser más fácil: tan sólo escríbase "*s* significa que *p*", e imagínese a "*p*" reemplazado por una oración. Las oraciones, como hemos visto, no pueden nombrar significados, y las oraciones con "que" prefijado no son nombres en absoluto, a menos que nosotros lo decidamos. Pero parece que surgen problemas por otro frente, pues es razonable esperar que en el forcejeo con la lógica del aparentemente no extensional "significa que" encontraremos problemas tan difíciles como, si no idénticos a, los problemas que nuestra teoría intenta resolver.

La única manera de tratar con esta dificultad que yo conozco es simple y extrema. Nos brota la inquietud de estar atrapados en lo intensional por usar las palabras "significa que" como relleno entre la descripción de la oración y la oración, pero puede ser que el éxito de nuestra misión dependa no del relleno sino de lo que éste rellena. La teoría habrá cumplido su tarea si proporciona, para cada oración *s* del lenguaje de estudio, una oración compañera (para reemplazar a "*p*") que, de algún modo todavía por aclarar, "dé el significado" de *s*. Un candidato obvio a oración compañera es justamente la propia *s*, si el lenguaje objeto está contenido en el metalenguaje; si no, una traducción de *s* en el metalenguaje. Como un audaz último paso, intentemos tratar extensionalmente la posición ocupada por "*p*": para implementar esto, barramos al oscuro "significa que", proveyamos a la oración que reemplaza a "*p*" con un conectivo oracional apropiado, y suministremos su propio predicado a la descripción que reemplaza a "*s*". El resultado plausible es:

(T) *s* es *T* si y sólo si *p*.

Lo que requerimos de una teoría del significado para un lenguaje *L* es que, sin apelar a (más) nociones semánticas, aplique suficientes restricciones al predicado "es *T*" como para implicar a todas las oraciones obtenidas del esquema *T* cuando "*s*" es reemplazada por una descripción estructural de una oración de *L* y "*p*" es reemplazada por esa oración.

Dos predicados cualesquiera que satisfagan esta condición tienen

la misma extensión, [6] luego si el metalenguaje es lo suficientemente rico, nada impide poner lo que yo llamo una teoría del significado bajo la forma de una definición explícita de un predicado "es *T*". Pero estén definidas explícitamente o caracterizadas recursivamente, es claro que las oraciones a las que se aplica el predicado "es *T*" serán justamente las oraciones verdaderas de *L*, puesto que la condición que hemos requerido para teorías del significado satisfactorias es en esencia la Convención *T* de Tarski, que pone a prueba la adecuación de una definición semántica formal de verdad.[7]

Hemos recorrido un tortuoso sendero para llegar hasta este punto, pero la conclusión puede establecerse con sencillez: una teoría del significado para un lenguaje *L* muestra "cómo los significados de las oraciones dependen de los significados de las palabras" si contiene una definición (recursiva) de verdad para *L*. Y, al menos hasta ahora, no tenemos otras ideas para resolver el problema. Vale la pena destacar que el concepto de verdad no jugó ningún papel ostensible cuando enunciamos nuestro problema original. Ese problema, una vez pulido, nos llevó a ver que una teoría del significado adecuada debe caracterizar a un predicado que reúna ciertas condiciones. Era propio de un descubrimiento el que tal predicado fuera exactamente aplicable a las oraciones verdaderas. Espero que cuanto estoy diciendo pueda describirse en parte como una defensa de la importancia filosófica del concepto semántico de verdad de Tarski. Pero mi defensa sólo se relaciona lejanamente, si es que lo hace, con la cuestión de si el concepto que Tarski mostró cómo definir es la (o una) concepción de verdad filosóficamente interesante, o con la cuestión de si Tarski ha echado alguna luz sobre el uso habitual de palabras como "verdadero" y "verdad". Es una lástima que el polvo de futiles y confusas batallas alrededor de estas cuestiones haya evitado que quienes tienen un interés teórico por el lenguaje —filósofos, lógicos, psicólogos y lingüistas— vieran en el concepto semántico de verdad (bajo el nombre que fuere) los cimientos sofisticados y poderosos de una teoría del significado competente.

No hay necesidad de suprimir, desde luego, la obvia conexión entre una definición de verdad del tipo que Tarski mostró cómo construir y el concepto de significado. La conexión es la siguiente: la definición funciona dando condiciones necesarias y suficientes para la verdad de cada oración, y dar condiciones de verdad es una forma de dar el significado de una oración. Conocer el concepto semántico de verdad para un lenguaje es conocer qué es para una oración —cualquier oración— ser verdadera, y esto equivale, en un buen sentido de la frase, a comprender el lenguaje. En todo caso, ésta es mi excusa por una característica de la presente discusión capaz de sacudir a viejos expertos: el libre uso que

[6] Suponiendo, desde luego, que la extensión de estos predicados está limitada a las oraciones de *L*.

[7] A. Tarski, "The Concept of Truth in Formalized Languajes".

doy a la palabra "significado", por lo cual lo que llamo teoría del significado terminó después de todo por no dar uso alguno a los significados, tanto de oraciones como de palabras. Ciertamente, dado que una definición de verdad del tipo de la de Tarski provee todo cuanto hemos exigido hasta ahora de una teoría del significado, resulta claro que una teoría así entra cómodamente dentro de lo que Quine denomina "teoría de referencia", que se distingue de lo que él denomina "teoría del significado". Hay bastante en favor de lo que yo llamo una teoría del significado, y bastante, quizás, en contra.[8]

Una teoría del significado (en el sentido tibiamente perverso que yo le doy) es una teoría empírica, y su ambición es dar cuenta del funcionamiento de un lenguaje natural. Como toda teoría, se la puede probar comparando algunas de sus consecuencias con los hechos. En el presente caso esto es fácil, pues la teoría opera en un infinito fluir de oraciones, cada una de las cuales da las condiciones de verdad de una oración; sólo necesitamos averiguar, en los casos estudiados, si lo que la teoría declara como condiciones de verdad de una oración realmente lo son. Un caso de prueba típico podría ser decidir si la oración "La nieve es blanca" *es* verdadera si y sólo si la nieve es blanca. No todos los casos serán tan simples (por razones que esbozaremos), pero es evidente que este tipo de test no invita a hacer cálculos de aplicación muy optimistas. Una concepción aguda de lo que constituye una teoría en este dominio proporciona un contexto estimulante para elaborar cuestionamientos profundos acerca de cuándo es correcta una teoría del lenguaje y cómo se la debe poner a prueba. Pero las dificultades son teóricas, no prácticas. En cuanto a la aplicación, el problema es obtener una teoría que se aproxime a la operatividad; cualquiera puede decir si es correcta.[9] Se puede ver por qué esto es así. La teoría no dice nada nuevo acerca de las condiciones bajo las cuales una oración individual es verdadera; no hace aparecer a esas condiciones más claramente que lo que la oración misma lo hace. La tarea de la teoría es relacionar las condiciones de verdad conocidas de cada oración con aquellos aspectos ("palabras") de la oración que se repiten en otras oraciones, y a las que pueden asignarse idénticos papeles en otras oraciones. El poder empírico de una teoría de este tipo depende de su éxito en recuperar la estructura de una capacidad muy complicada: la capacidad de hablar y comprender un lenguaje.

[8] Pero puede citarse a Quine a favor del uso que yo propongo: "...en cuanto al *significado*... puede decirse que una palabra está determinada según el grado en que la verdad o falsedad de sus contextos está determinado". (*"Truth by Convention"*, 82). Puesto que una definición de la verdad determina el valor de verdad de cada oración en el lenguaje objeto (relativo a una oración en el metalenguaje), determina el significado de cada palabra y oración. Esto parecería justificar el título de Teoría del Significado.

[9] Para dar un único ejemplo: el que una teoría implique que " 'La nieve es blanca' es verdadero si y sólo si la nieve es blanca" es claramente un punto a favor de ella. Pero idear una teoría que implique esto (y funcione para todas las oraciones relacionadas) no es tarea trivial. No conozco una teoría completamente satisfactoria que tenga éxito con este mismo caso (el problema de los "términos masa").

Podemos determinar con bastante facilidad si los pronunciamientos particulares de la teoría concuerdan con nuestra comprensión del lenguaje; esto es consistente con una débil percepción del diseño de la maquinaria de nuestros logros lingüísticos.

Las afirmaciones del último párrafo sólo son de aplicación directa para el caso especial en que se supone que el lenguaje para el cual se está caracterizando la verdad es parte del lenguaje que usa y comprende el caracterizador. Bajo estas circunstancias, el constructor de una teoría se valdrá cuando pueda, como cosa natural, de la conveniencia propia de un metalenguaje al tratar una oración que se garantiza equivalente a cada oración del lenguaje objeto. No obstante, este hecho no debería confundirnos y hacernos pensar que una teoría que incluye a " 'La nieve es blanca' es verdadero si y sólo si la nieve es blanca" es más correcta que una que en cambio incluye a (*S*) "La nieve es blanca" es verdadero si y sólo si el césped es verde, descontando, desde luego, que estamos tan seguros de la verdad de (*S*) como lo estamos de la verdad de su más celebrada predecesora. Pero (*S*) puede no alentar la misma confianza de que una teoría que la implique merezca denominarse una teoría del significado.

La amenaza de una mengua en el vigor de la teoría se puede contrarrestar como sigue. Lo grotesco de (*S*) en sí mismo no significa nada en contra de una teoría de la cual es consecuencia, siempre y cuando la teoría dé los resultados correctos para cada oración (en base a su estructura, no habiendo otro modo). No es fácil de ver cómo (*S*) podría ser partícipe de un emprendimiento tal, pero si lo fuera —esto es, si (*S*) resultara de una caracterización del predicado "es verdadero" que hubiera llevado a un invariable apareamiento de verdades con verdades y de falsedades con falsedades— entonces, pienso yo, no habría nada esencial a la idea de significado que restara capturar.[10]

Lo que aparece a la derecha del bicondicional en oraciones de la forma "*s* es verdadera si y sólo si *p*" cuando dichas oraciones son consecuencias de una teoría de la verdad juega su papel en la determinación del significado de *s* no pretendiendo una sinonimia sino agregando una pincelada más a la imagen que, tomada como un todo, nos dice qué es posible saber del significado de *s*; la pincelada se puede agregar en virtud del hecho de que la oración que reemplaza a "*p*" es verdadera si y sólo si *s* lo es.

Puede ayudar a reflejar que (*S*) sea aceptable, si lo es, el que estemos independientemente seguros de la verdad de "La nieve es blanca" y de "El césped es verde"; pero en casos en que no estamos seguros de la verdad de una oración, podemos tener confianza en una caracterización

[10] A menudo los críticos han pasado por alto inadvertidamente la condición esencial que se menciona en este párrafo. El punto es que (*S*) no podría pertenecer a ninguna teoría razonablemente simple que también diera las condiciones de verdad correctas para "Eso es nieve" y "Esto es blanco". (Véase la discusión de las expresiones deícticas más adelante.) (Nota agregada en 1982.)

del predicado de verdad sólo si se aparea esa oración con otra de la que tenemos buenas razones para considerar equivalente. Sería un error que alguien que tuviera alguna duda respecto del color de la nieve o del césped aceptara una teoría que produjera a (*S*), aun si sus dudas fueran de igual grado, a menos que pensara que el color de la una estuviera supeditado al color del otro.[11] La omnisciencia puede obviamente deparar más teorías del significado fantásticas que la ignorancia; por otra parte, la omnisciencia tiene menos necesidad de comunicación.

Por supuesto, debe ser posible para el hablante de un lenguaje construir una teoría del significado para el hablante de otro, si bien en este caso el test empírico para determinar si la teoría es correcta no será ya trivial. Como antes, la meta de la teoría será una correlación infinita de oraciones similares en verdad. Pero no se deberá suponer esta vez que el constructor de la teoría tiene una percepción directa de las equivalencias probables entre su propia lengua y la extranjera. Lo que él debe hacer es descubrir, como quiera que pueda, las oraciones que el extranjero considera verdaderas en su propia lengua (o aun mejor, hasta qué punto las considera verdaderas). El lingüista intentará luego construir una caracterización de la verdad para el extranjero que produzca, en lo posible, un mapeo de las oraciones consideradas verdaderas (o falsas) por el extranjero con las oraciones consideradas verdaderas (o falsas) por el lingüista. Suponiendo que no se encuentra una coincidencia perfecta, el residuo de oraciones consideradas verdaderas traducidas por oraciones consideradas falsas (o viceversa) es el margen de error (foráneo o doméstico). Cierta caridad en la interpretación de las palabras y los pensamientos de los otros es inevitable también en otra dirección: así como debemos maximizar el acuerdo o arriesgarnos a no encontrar sentido a lo que el extranjero está diciendo, también debemos maximizar la autoconsistencia que le atribuimos, bajo pena de no entenderlo *a él*. No emerge un único principio de caridad óptima; así, las restricciones no determinan una única teoría. En una teoría de la traducción radical (como la llama Quine) no se desenmarañan completamente las cuestiones de lo que el extranjero significa de las cuestiones de lo que él cree. No sabemos lo que alguien significa a menos que sepamos lo que cree; no sabemos lo que alguien cree a menos que sepamos lo que significa. En la interpretación radical estamos capacitados para entrar en este círculo, aunque más no sea en forma incompleta, porque a veces podemos decir que una persona accede a una oración que nosotros no comprendemos.[12]

[11] Este párrafo es confuso. Debería decir que las oraciones de la teoría son generalizaciones empíricas acerca de hablantes, y por lo tanto deben ser no sólo verdaderas sino también lícitas. (*S*) presumiblemente no es una ley, pues no sustenta contrafácticos apropiados. También es de notar que la evidencia para aceptar las condiciones de verdad (relativizadas a tiempos y hablantes) para "Eso es nieve" se basa en la conexión causal entre el asentimiento de un hablante frente a la oración y la presentación demostrativa de la nieve. Para mayor discusión véase el Ensayo 12. (Nota agregada en 1982.)

[12] Este esquema de la puesta a prueba de una teoría de significado para una lengua extranjera obviamente debe su inspiración a la descripción que Quine hace de la traduc-

En estas últimas páginas he inquirido acerca de cómo puede ser puesta a prueba empíricamente una teoría del significado que toma la forma de una definición de verdad, y he ignorado alegremente la cuestión crucial de si hay alguna posibilidad seria de proveer una teoría de ese tipo para un lenguaje natural. ¿Cuáles son las perspectivas para una teoría semántica formal de un lenguaje natural? Muy pobres, según Tarski; y creo que la mayor parte de los lógicos, filósofos del lenguaje y lingüistas concuerda.[13] Déjenme hacer lo que pueda por disipar ese pesimismo. Lo que pueda en una manera general y programática, desde luego, pues la clasificación final dependerá ciertamente de la prueba de los teoremas correctos.

Tarski concluye la primera sección de su clásico ensayo sobre el concepto de verdad en los lenguajes formalizados con las siguientes afirmaciones, que pone en bastardilla:

> ...La posibilidad misma de un uso consistente de la expresión "oración verdadera" que esté en armonía con las leyes de la lógica y con el espíritu del lenguaje diario parece ser muy cuestionable, y consecuentemente la misma duda se aplica a la posibilidad de construir una definición correcta de esta expresión. (165)

Más adelante, en el mismo ensayo, vuelve sobre el asunto:

> ...cuando el concepto de verdad (al igual que otros conceptos semánticos) se aplica al lenguaje coloquial en conjunción con las leyes normales de la lógica, lleva inevitablemente a confusiones y contradicciones. Quien desee, a pesar de todas las dificultades, perseguir el dominio de la semántica del lenguaje coloquial con el auxilio de métodos exactos, se verá conducido primero a encarar la ingrata tarea de una reforma de este lenguaje. Encontrará necesario definir su estructura, superar la ambigüedad de los términos que en él aparecen, y finalmente dividir el lenguaje en una serie de lenguajes de mayor y mayor extensión, cada uno de los cuales estará vinculado con el siguiente en la misma relación con la cual un lenguaje formalizado se vincula con su metalenguaje. Sin embargo, se puede dudar de que el lenguaje de la vida diaria, tras ser "racionalizado" de este modo, conserve su naturalidad y no tome en cambio las características de los lenguajes formalizados. (267)

Aquí emergen dos temas: que el carácter universal de los lenguajes naturales lleva a contradicción (las paradojas semánticas), y que los

---

ción radical en el capítulo II de *Word and Object*. Voy más lejos que Quine cuando sugiero que una teoría de la traducción radical aceptable toma la forma de una caracterización recursiva de la verdad. Hacia el final de este trabajo, en la discusión de los demostrativos, surgirá otra fuerte coincidencia con Quine.

[13] Por lo que yo sé, se ha discutido muy poco acerca de si se puede dar una definición formal de verdad para un lenguaje natural. Pero en términos más generales, varios autores han insistido en la aplicación de los conceptos de la semántica formal al lenguaje natural. Véanse por ejemplo, las contribuciones de Yehoshua Bar-Hillel y de Evert Beth en *The Philosophy of Rudolph Carnap*, y *"Logical Syntax and Semantics"* de Bar-Hillel.

lenguajes naturales son demasiado confusos y amorfos para permitir la aplicación directa de métodos formales. El primer punto merece una respuesta seria, y yo quisiera tenerla. Como están las cosas, sólo diré por qué pienso que podemos justificar el seguir adelante sin haber desinfectado esta fuente particular de inquietud conceptual. La paradojas semánticas surgen cuando el rango de los cuantificadores en el lenguaje objeto es demasiado generoso en ciertos aspectos. Pero no es del todo claro cuán injusto sería para con el urdu o el vasco considerar al rango de sus cuantificadores insuficiente para ofrecer una definición explícita de "verdadero en urdu" o "verdadero en vasco". O, para decirlo de otra manera, quizá más seria, siempre puede haber, dada la naturaleza del caso, algo que captamos al intentar comprender el lenguaje del otro (el concepto de verdad) pero que no podemos comunicárselo. En todo caso, la mayor parte de los problemas de interés filosófico general surgen dentro de un fragmento del lenguaje natural relevante que, consideramos, contiene muy poco de teoría de conjuntos. Desde luego que estos comentarios no cumplen con la pretensión de que los lenguajes naturales son universales. Pero me parece que esta pretensión, ahora que sabemos que tal universalidad lleva a paradojas, es sospechosa.

El segundo punto del texto de Tarski sugiere que tendríamos que reformar un lenguaje natural hasta dejarlo irreconocible antes de poder aplicar métodos semánticos formales. Si esto es cierto, resulta fatal para mi proyecto, pues la tarea de una teoría del significado tal como yo la concibo no es cambiar, mejorar o reformar un lenguaje, sino describirlo y comprenderlo. Miremos el lado positivo de esto. Tarski ha mostrado la manera de producir una teoría para diferentes tipos de lenguajes formales interpretados; tómese uno tan parecido al castellano como sea posible. Puesto que este nuevo lenguaje ha sido explicado en castellano y contiene mucho castellano no sólo podemos, sino que pienso que debemos, considerarlo parte del castellano de aquellos que lo comprenden. Para este fragmento del castellano tenemos, *ex hypothesi*, una teoría del tipo requerido. No sólo eso, sino que al interpretar a este adjunto del castellano en el viejo castellano necesariamente insinuamos conexiones entre el viejo y el nuevo. Como quiera que haya oraciones del viejo castellano con las mismas condiciones de verdad de oraciones del adjunto, podemos extender la teoría para incluirlas. Lo que se exige es ante todo mecanizar cuanto sea posible lo que hoy hacemos merced a una destreza cuando expresamos el castellano común en una u otra notación canónica. El asunto no es que la notación canónica sea mejor que la tosca forma idiomática original, sino más bien que si nosotros sabemos *de cuál* expresión idiomática la notación canónica es canónica, tenemos una teoría tan buena para la forma idiomática como para su compañera.

Los filósofos llevan mucho tiempo abocados a la dura tarea de aplicar la teoría al lenguaje ordinario mediante el recurso de aparear oraciones del lenguaje vernáculo con oraciones para las cuales tienen una teoría. Frege hizo una importantísima contribución al mostrar de qué manera puede dominarse, en algunos de sus usos, a "todos", "algunos", "algún", "todo", "cada", "ningún" y otros pronombres asociados; por pri-

mera vez fue posible soñar con una semántica formal para una parte significativa de un lenguaje natural. Este sueño se convirtió claramente en realidad gracias a la obra de Tarski. Sería una injusticia no destacar que gracias a los espléndidos logros de Frege y de Tarski hemos conseguido penetrar profundamente en la estuctura de nuestras lenguas maternas. Los filósofos de tendencia lógica se han inclinado por comenzar desde donde había llegado la teoría para desde allí labrarse un camino hacia las complicaciones del lenguaje natural. Los lingüistas contemporáneos, tras objetivos que no se pueden distinguir fácilmente de los de aquéllos, arrancan de lo ordinario y trabajan en busca de una teoría general. Si alguna de las dos partes tiene éxito, debe producirse un encuentro. El trabajo reciente de Chomsky y otros está haciendo mucho por someter las complejidades de los lenguajes naturales bajo una teoría seria. Para dar un ejemplo: supongamos que se determinan con éxito las condiciones de verdad para un rango significativo de oraciones en voz activa. Luego, merced a un procedimiento formal de transformación de cada una de estas oraciones en una oración correspondiente de la voz pasiva, la teoría de la verdad puede extenderse de manera obvia a este nuevo conjunto de oraciones.[14]

Hay un problema que Tarski roza al pasar y que no necesita ser resuelto, al menos en todas sus manifestaciones, para llevar adelante la teoría: la existencia de "términos ambiguos" en los lenguajes naturales. Hasta tanto la ambigüedad no afecte la forma gramática, y pueda traducirse, ambigüedad por ambigüedad, al metalenguaje, una definición de verdad no nos transmitirá ninguna mentira. La principal dificultad que la frase *"believes that"* en inglés presenta para la semántica sistemática no reside en su vaguedad, su ambigüedad o su incompetencia para incorporarla a una ciencia seria: tomemos al inglés como metalenguaje, y todos *estos* problemas se trasladarán sin más al metalenguaje. Pero el problema central de la gramática lógica de *"believes that"* nos seguirá merodeando.

El ejemplo es apto para ilustrar otra cuestión relacionada con aquélla, pues la discusión de las oraciones de creencia ha pasado casi

[14] El *rapprochement* que anticipadamente imagino entre la gramática tranformacional y una lúcida teoría del significado se ha visto muy favorecido por un cambio reciente en el concepto de gramática transformacional que describió Chomsky en el artículo antes citado (nota 5). Las estructuras que genera la parte de estructura de frases de la gramática, tal como se ha comprometido últimamente, son aquellas pasibles de interpretación semántica; pero esta visión es inconsistente con la idea, que hasta hace poco sostuvo Chomsky, de que las operaciones recursivas se introducen sólo mediante las reglas de transformación. Chomsky cree hoy que las reglas de estructura de frase son recursivas. Puesto que los lenguajes a los cuales se aplican directa y naturalmente los métodos semánticos formales son aquellos para los cuales resulta apropiada una gramática de estructura de frase (recursiva), resulta claro que la visión actual de Chomsky de la relación entre las estructuras generadas por la parte de estructura de frase de la gramática y las oraciones del lenguaje es muy parecido a la visión que tienen muchos lógicos y filósofos de la relación entre los lenguajes formalizados, más ricos, y el lenguaje ordinario. (Debo mucho a Bruce Vermazen por estas observaciones.)

siempre por alto una distinción fundamental de sus tareas: develar la gramática o forma lógica de las oraciones (lo cual es de incumbencia de una teoría del significado tal como yo la concibo), y el análisis de palabras o expresiones individuales (que la teoría considera primitivos). Así Carnap, en la primera edición de *Meaning and Necessity*, sugirió que transformáramos "John cree *that* la Tierra es redonda" en "John responde afirmativamente a 'la Tierra es redonda' como oración del inglés". Descartó esto cuando Mates señaló que John podría responder afirmativamente a una oración y no a otra, no importa cuán cercanos fueran sus significados.[15] Pero aquí hay una confusión desde el punto de partida. La estructura semántica de una oración de creencia, de acuerdo con esta idea de Carnap, está dada por un predicado triádico con espacios reservados para expresiones que refieren a una persona, una oración, y un lenguaje. Intentar un análisis de este predicado, quizá según criterios conductistas, es un tipo de problema completamente diferente. De entre los méritos de la concepción de Tarski de una teoría de la verdad, uno de los que se destaca es que la pureza de método que nos exige resulta de la formulación del problema mismo y no de la restricción autoimpuesta de algún puritanismo filosófico adventicio.

Creo que no es posible exagerar las ventajas que representan para la filosofía del lenguaje el tener presente la distinción entre cuestiones de forma o gramática lógicas, y el análisis de conceptos individuales. Otro ejemplo puede ayudar a explicitarlo.

Si suponemos que las cuestiones de gramática lógica están resueltas, oraciones como "Bardot es buena" no generan problemas especiales para una definición de verdad. Las profundas diferencias entre términos descriptivos y evaluativos (emotivos, expresivos, etcétera) no se manifiestan aquí. Incluso si sostenemos que en algún sentido importante las oraciones morales o evaluativas carecen de un valor de verdad (por ejemplo, porque no pueden verificarse), no debemos retroceder ante " 'Bardot es buena' es verdadero si y sólo si Bardot es buena"; en una teoría de la verdad, esta consecuencia debe acompañar al resto, siguiendo la pista, como debe ser, de la situación semántica de tales oraciones en la totalidad del lenguaje (de su relación con las generalizaciones, de su papel en oraciones compuestas como "Bardot es buena y Bardot es tonta", y así sucesivamente). Lo que es privativo de las palabras evaluativas aquí simplemente no se toca: el misterio se transfiere de la palabra "buena" en el lenguaje objeto a su traducción en el metalenguaje.

Pero "buena", tal como aparece en "Bardot es una buena actriz", es otra cosa. El problema no es que la traducción de esta oración no esté en el metalenguaje; supongamos que lo está. El problema es construir una definición de verdad tal que " 'Bardot es una buena actriz' es verdadero si y sólo si Bardot es una buena actriz" —y todas las otras oraciones como ésta— sean consecuencias. Obviamente "buena actriz" no significa

[15] B. Mates, "Synonymity".

"buena y actriz". Podríamos pensar en tomar a "es una buena actriz" como un predicado no analizado. Esto borraría toda conexión entre "es una buena actriz" y "es una buena madre", y no nos daría excusas para pensar que "buena", en estos usos, es una palabra o elemento semántico. Y peor aun, nos impediría construir cualquier definición de verdad, pues no hay límite a la cantidad de predicados que tendríamos que tratar como lógicamente simples (y por lo tanto situar en cláusulas separadas en la definición de satisfacción): "es una buena compañera para los perros", "es una buena interlocutora de 28 años de edad", etcétera. El problema no es peculiar de este caso: es el problema de los adjetivos atributivos en general.

Es consistente con la actitud aquí tomada juzgar usualmente como un error estratégico el acometer un análisis filosófico de palabras o expresiones que no esté precedido o en todo caso acompañado por el intento de obtener la gramática lógica correcta. Pues, ¿cómo podemos confiar en nuestros análisis de palabras como "correcto", "debe", "puede" y "obligado", o de las frases que empleamos para hablar de acciones, sucesos y causas, cuando no sabemos con qué partes (lógicas, semánticas) del discurso tenemos que tratar? Diría algo muy parecido respecto de los estudios de la "lógica" de éstas y otras palabras, y de las oraciones que las contienen. No puede saberse si el esfuerzo y la inventiva aplicados al estudio de las lógicas deónticas, modales, imperativas y erotéticas han sido en gran medida futiles o no hasta tanto tengamos análisis semánticos aceptables de las oraciones que tales sistemas pretenden tratar. A veces, los filósofos y los lógicos hablan y trabajan como si fueran libres de elegir entre, digamos, el condicional veritativo funcional y otros, o libres de introducir operadores oracionales no veritativo-funcionales como "Sea el caso que" o "Tiene que ser el caso que". Pero de hecho la decisión es crucial. Cuando nos apartamos de expresiones idiomáticas que podemos acomodar en una definición de la verdad, nos deslizamos hacia (o creamos) un lenguaje para el cual no tenemos una versión semántica coherente; es decir, no tenemos ninguna explicación acerca de cómo ese discurso puede ser integrado dentro de lenguaje como un todo.

Volviendo a nuestro tema principal: hemos reconocido que una teoría del tipo propuesto deja a toda la cuestión del significado de las palabras individuales exactamente donde estaba. Aun cuando el metalenguaje es diferente del lenguaje objeto, la teoría no ejerce una presión tendiente a la mejora, clarificación o análisis de palabras individuales, excepto cuando, por accidente del vocabulario, la traducción directa fracasa. Así como en general no se trata la sinonimia entre expresiones, lo mismo ocurre con la sinonimia de oraciones y con la analiticidad. Incluso oraciones tales como "Una yegua es un caballo hembra" carecen de un rótulo especial a menos que deseemos proporcionárselo. Una definición de verdad no distingue entre oraciones analíticas y otras, a excepción de oraciones que deben su verdad a la sola presencia de las constantes que otorgan a la teoría su control sobre la estructura: la teoría no sólo implica que estas oraciones sean verdaderas sino que también seguirán siendo verdaderas bajo cualquier reescritura significativa de sus

partes no lógicas. Dada la aplicación limitada conferida a la noción de verdad lógica, las nociones asociadas de equivalencia lógica e implicación seguirán por igual camino. Es difícil imaginar cómo una teoría del significado podría fracasar hasta este extremo en leer una lógica dentro de su lenguaje objeto; y en la medida en que lo haga, nuestras intuiciones de verdad lógica, equivalencia e implicación pueden ser requeridas para la construcción y puesta a prueba de la teoría.

Me referiré ahora a otro asunto sumamente molesto: el hecho de que una misma oración dicha por una persona o en determinado momento puede ser verdadera y dicha por otra persona o en otro momento puede ser falsa. Tanto los lógicos como los críticos de los métodos formales parecen estar aquí mayormente (aunque de ninguna manera universalmente) de acuerdo en que la semántica formal y la lógica son incapaces de habérselas con las perturbaciones que causan los demostrativos. Los lógicos han reaccionado a menudo despreciando al lenguaje natural y tratando de mostrar cómo arreglárselas sin demostrativos; sus críticos reaccionan despreciando a la lógica y a la semántica formal. Nada de esto puede alegrarme: resulta claro que los demostrativos no pueden eliminarse de un lenguaje natural sin un perjuicio o un cambio radical, por lo cual no hay otra alternativa que acomodar la teoría a ellos. No surgen errores lógicos si simplemente tratamos a los demostrativos como constantes;[16] tampoco aparecen problemas si damos una definición de verdad semántica. " 'Yo soy sensato' es verdadero si y sólo si yo soy sensato", con su torpe indiferencia ante el elemento demostrativo en "Yo", sale de la línea de montaje junto con " 'Sócrates es sensato' es verdadero si y sólo si Sócrates es sensato" con *su* torpe indiferencia ante el elemento demostrativo en "es sensato" (el tiempo).

Con este tratamiento de los demostrativos se resiente no la definición de un predicado de verdad, sino la plausibilidad de la pretensión de que lo que se ha definido es la verdad. Pues esta pretensión es aceptable solamente si el hablante y las circunstancias de emisión de cada oración mencionados en la definición son equiparables al hablante y las circunstancias de emisión de la propia definición de verdad. Bien podría señalarse además que forma parte de la comprensión de los demostrativos el conocer las reglas según las cuales ellos adaptan su referencia a la circunstancia; si asimilamos los demostrativos a los términos constantes esta característica queda reducida a la nada. Podemos hacer frente a estas observaciones, creo yo, aunque sólo mediante una revisión de grandes proporciones en la teoría de la verdad. Meramente sugeriré cómo se podría hacer esto, si bien lo único que hace falta es una mera sugerencia: la idea es técnicamente trivial, y acorde con el trabajo que se está llevando a cabo en la lógica de los tiempos.[17]

Podemos tomar a la verdad como una propiedad, no de las oracio-

[16] Véase W. V. Quine, *Methods of Logic*, 8.

[17] Esta idea ha resultado ser ingenuamente optimista. Véase como obra seria sobre el tema S. Weinstein, "Truth and Demonstratives". (Nota agregada en 1982.)

nes, sino de las emisiones, o actos de habla, o tríos ordenados de oraciones, tiempos y personas; pero es más simple ver a la verdad sólo como una relación entre una oración, una persona y un tiempo. Bajo este tratamiento, la lógica común tal como ahora se la entiende se puede aplicar de manera usual, pero solamente a conjuntos de oraciones relativizadas al mismo hablante y tiempo; otras relaciones lógicas entre oraciones dichas en diferentes tiempos y por diferentes hablantes pueden articularse mediante nuevos axiomas. Eso no me preocupa. La teoría del significado sufre un cambio sistemático pero no confuso; a cada expresión con un elemento demostrativo debe corresponder en la teoría una frase que relacione las condiciones de verdad de oraciones en las cuales se presenta la expresión con los hablantes y tiempo cambiantes. Así, la teoría implicará oraciones como la siguiente: "Yo estoy cansado" es verdadera cuando es (potencialmente) dicha por *p* en *t* si y sólo si *p* está cansado en *t*.

"Robaron ese libro" es verdadera cuando es (potencialmente) dicha por *p* en *t* si y sólo si el libro señalado por *p* en *t* fue robado antes de *t*.[18]

Evidentemente, este camino no muestra cómo eliminar a los demostrativos; por ejemplo, no se sugiere que "el libro señalado por el hablante" pueda sustituirse ubicuamente por "ese libro" *salva veritate*. El hecho de que los demostrativos estén sujetos a tratamiento formal debe alentar en gran medida las esperanzas de obtener una semántica seria del lenguaje natural, pues es posible que muchos enigmas notorios, como el análisis de citas o de oraciones acerca de actitudes proposicionales, puedan resolverse si reconocemos una construcción demostrativa oculta.

Ahora que hemos relativizado la verdad a los tiempos y a los hablantes, es apropiado echar otra mirada al problema de poner a prueba empíricamente a una teoría del significado para una lengua extranjera. La esencia del método era, recordemos, correlacionar oraciones consideradas verdaderas con oraciones consideradas verdaderas por medio de una definición de verdad, y dentro de los límites del error inteligible. Ahora debe elaborarse el cuadro para dar cabida al hecho de que las oraciones son verdaderas, y consideradas verdaderas, sólo en relación a un hablante y a un tiempo. Es obvio que las oraciones con demostrativos proveen un test muy sensible de la corrección de una teoría del significado, y constituyen la ligazón más directa entre el lenguaje y los objetos macroscópicos recurrentes de interés y atención humanos.[19]

En este trabajo he supuesto que los hablantes de un lenguaje pueden determinar efectivamente el significado o los significados de una ex-

[18] Hay algo más que un indicio de esta aproximación a los demostrativos y la verdad en J. L. Austin, "Truth".

[19] Estas observaciones derivan de la idea de Quine de que las "oraciones de ocasión" (aquéllas con un elemento demostrativo) deben desempeñar un papel central en la construcción de un manual de traducción.

presión arbitraria (si es que tiene un significado), y que la tarea principal de una teoría del significado es mostrar cómo es posible esto. He sostenido que una caracterización de un predicado de verdad describe el tipo de estructura requerido y proporciona un criterio claro y comprobable para una semántica adecuada a un lenguaje natural. Sin duda, hay otros requerimientos que razonablemente pueden exigirse a una teoría del significado. Pero una teoría que no hace otra cosa que definir la verdad para un lenguaje está mucho más próxima a constituir una teoría del significado completa que lo que un análisis superficial podría sugerir; esto es, al menos, lo que yo he afirmado.

Dado que a mi parecer no hay otra alternativa, me he hecho una imagen optimista y programática de las posibilidades de una caracterización formal de un predicado de verdad para un lenguaje natural. Pero debe admitirse que nos espera una apabullante lista de dificultades y acertijos. Para nombrar a unos pocos: no conocemos la forma lógica de las oraciones contrafácticas o subjuntivas; tampoco la de las oraciones acerca de probabilidades y acerca de relaciones causales; no tenemos una buena idea del rol lógico de los adverbios, ni del de los adjetivos atributivos; no tenemos una teoría para los términos masa como "fuego", "agua" y "nieve", ni para las oraciones de creencia, percepción e intención, ni para verbos de acción que implican propósito. Y finalmente, están todas las oraciones que no parecen tener valor de verdad alguno: las imperativas, optativas, interrogativas y muchas otras. Una amplia teoría del significado para un lenguaje natural debe enfrentar con éxito cada uno de estos problemas.[20]

[20] Se encontrarán soluciones tentativas a algunos de estos problemas en los Ensayos 6 al 10 de *Essays on Actions and Events*, y en los Ensayos 6 al 8 de este libro. También se trata el tema de los Ensayos 3, 4, 9 y 10, y se hace referencia a algunos progresos en la sección 1 del Ensayo 9.

3

Fiel a los hechos

Un enunciado verdadero es un enunciado fiel a los hechos. Esta observación parece incorporar el mismo tipo de juicio obvio y esencial acerca de la verdad como el siguiente acerca de la maternidad: una madre es una persona que es madre de alguien. La *propiedad* de ser una madre está explicada por la *relación* entre una mujer y su hijo; de manera similar, esto parece sugerir que la propiedad de ser verdadero será explicada por una relación entre un enunciado y alguna otra cosa. Sin prejuzgar la cuestión de qué podría ser esa otra cosa, o de qué palabra o frase expresa mejor la relación (de ser verdadero para corresponder al ilustrar), me tomaré la licencia de llamar *teoría de correspondencia* de la verdad a cualquier consideración de este tipo.

Las teorías de correspondencia descansan sobre lo que parece ser una idea inevitable pero simple, aunque no han respondido bien al examen. La dificultad principal está en encontrar una noción de hecho que explique todo, que no caiga, al estudiarla, en lo trivial o lo vacío. Así, la discusión reciente se ocupa fundamentalmente de decidir si alguna forma de correspondencia es verdadera y trivial ("...la teoría de la verdad es una serie de verdades de Perogrullo")[1] o, en tanto no sea confusa, simplemente vacía ("La teoría de correspondencia no requiere purificación, sino eliminación").[2] Quienes han discutido el concepto semántico de verdad en conexión con las teorías de correspondencia, por lo general han catalogado al concepto semántico como irrelevante o trivial.

En este trabajo defiendo una versión de la teoría de correspondencia. Pienso que la verdad puede explicarse apelando a una relación entre el lenguaje y el mundo, y que el análisis de esa relación ofrece una idea de cómo, al expresar oraciones, a veces logramos discernir lo que es verdadero. El concepto semántico de verdad, que Tarski expuso sistemáticamente por primera vez,[3] desempeñará un papel esencial en esta defensa. Habría posibilidades de probar que toda teoría o definición de la verdad que cumpla normas plausibles contiene necesariamente recursos conceptuales adecuados para definir un sentido de correspondencia. Mi proyecto es menos ambicioso: me sentiré satisfecho si puedo encontrar

[1,2] J. L. Austin y P. F. Strawson, simposio sobre "La Verdad". Las observaciones citadas son de Austin y de Strawson respectivamente, y aparecen en las páginas 115 y 129.
[3] A. Tarski, "The Concept of Truth in Formalized Languages".

una interpretación natural de la relación de correspondencia que ayude a explicar la verdad. Sin duda, sería consistente con el éxito de este intento hallar una fórmula para eliminar frases como "es verdad que" y "es verdadero" de muchos o todos los contextos: teorías de correspondencia y de redundancia no necesariamente entran en conflicto. No obstante, podemos hallar indicios acerca del rol de la correspondencia estudiando los resultados que obtendríamos al reemplazar sistemáticamente oraciones con palabras o frases veritativas por oraciones sin ellas.

La oración

(1) El enunciado de que el francés es la lengua oficial de Mauricio es verdadero.

es materia equivalente a "El francés es la lengua oficial de Mauricio"; y lo mismo podría decirse de dos oraciones cualesquiera relacionadas de manera similar. Esto impulsa la idea de que las palabras que dan asiento a la oración anidada en (1) representan una función veritativa de identidad, con igual poder que la doble negación, pero carente de articulación significativa. Según esta sugerencia, el que (1) consista en un término singular complejo y un predicado no sería más que un capricho de la gramática.

La teoría de doble negación de la verdad tiene el problema de que sólo es aplicable a oraciones que tienen oraciones anidadas, como (1) y "Es verdadero que 2 + 2 = 5". La teoría no puede hacer frente directamente a:

(2) El teorema de Pitágoras es verdadero.

(3) Nada que haya dicho Aristóteles es verdadero.

Podríamos sostener que la teoría de doble negación es aplicable a (1) en tanto que dejaríamos a (2) y a (3) para su tratamiento por separado. Pero es difícilmente plausible que las palabras "es verdadero" tengan diferentes significados en estos diferentes casos, sobre todo cuando parece haber simples inferencias que conectan oraciones de los dos tipos. Así, de (2) y de "El teorema de Pitágoras es el enunciado de que el cuadrado de la hipotenusa es igual a la suma de los cuadrados de los otros dos lados" podemos inferir "El enunciado de que el cuadrado de la hipotenusa es igual a la suma de los cuadrados de los lados es verdadero".

Es tentador pensar que la teoría de doble negación puede extenderse de algún modo para cubrir oraciones como (2) y (3). El razonamiento podría ser así: la teoría de doble negación nos dice que para cada enunciado hay una oración que lo expresa. Pero entonces (2) es válida para el caso de que haya una oración verdadera que exprese el teorema de Pitágoras, y (3) es válida en caso de que ninguna oración verdadera exprese algo que haya dicho Aristóteles. Se verá que la aparente necesidad de usar la palabra "verdadero" en esta explicación es inofensiva, al expresar (2) y (3) así:

(2') (p) (El enunciado de que p = el teorema de Pitágoras $\rightarrow p$)

(3') $\sim (\exists p)$ (Aristóteles dijo que $p.p$)

Estamos siguiendo ahora una línea que diverge de la simple teoría de doble negación por aceptar una ontología de enunciados, y por introducir cuantificación en posiciones que pueden ser ocupadas por oraciones. No es que las variables en (2') y en (3') abarquen los enunciados; se trata más bien de que expresiones como las que flanquean al signo de identidad en (2') refieren a enunciados. En la teoría de doble negación, la referencia putativa a enunciados y la predicación putativa de verdad fueron absorbidas en una expresión gramáticamente compleja pero lógicamente simple, un conectivo oracional veritativo funcional. En cambio, la teoría presente nos permite ver a "es verdadero" como un predicado genuino. Ella nos proporciona el siguiente principio:

(4) (p) (El enunciado de que p es verdadero $\leftrightarrow p$)

que lleva oraciones libres del predicado "es verdadero" y lógicamente equivalentes a oraciones que lo contienen. Aqui no se justifica la verdad como algo que puede predicarse de los enunciados, sino que se la explica.

Se la explica, claro, si comprendemos (2'), (3') y (4). Pero ¿las comprendemos? El problema está en las variables. Dado que las variables reemplazan a oraciones tanto cuando se presentan después de palabras como "Aristóteles dijo *that*" como en contextos veritativos funcionales; la extensión de las variables debe consistir en entidades que pueden interpretarse como nombradas por las oraciones, en ambos usos. Pero hay razones muy poderosas, como lo señaló Frege, para suponer que si las oraciones, cuando aparecen solas o en contextos veritativos funcionales, nombran algo, luego todas las oraciones verdaderas nombran la misma cosa.[4] Esto nos forzaría a concluir que el enunciado de que p es idéntico al enunciado de que q siempre y cuando p y q sean ambos verdaderos; presumiblemente un resultado inaceptable.

En un breve pasaje que se menciona a menudo, F. P. Ramsey presenta una teoría similar, si no idéntica a la que acabamos de discutir. El autor observa que oraciones como (2'), (3') y (4) no pueden leerse de modo convincente en inglés si no se agregan las palabras "es verdadero" al final, pero parece ver esto como un capricho, o incluso un defecto, del lenguaje (agregamos "es verdadero" pues olvidamos que "p" ya contiene un verbo "variable"). Ramsey luego dice:

> Quizás esto puede aclararse mejor suponiendo por un momento que sólo se cuestiona una forma de proposición, digamos, la forma relaciona aRb; luego "El siempre tiene razón" podría expresarse como "Para todo a, R, b, si él afirma aRb, luego aRb", para la cual "es verdadero" sería un agregado obviamente superfluo. Cuando se incluye a todas las formas de proposiciones el análisis es más complicado pero no esencialmente diferente.[5]

[4] G. Frege, "On Sense and Reference". Por el argumento, véase el Ensayo 2.
[5] F. P. Ramsey, "Facts and Propositions", 143.

Pienso que debemos suponer que Ramsey quiere que las variables "*a*" y "*b*" abarquen individuos de algún tipo, y que "*R*" abarque relaciones (diádicas). Entonces su versión de "El siempre tiene razón" estaría mejor expresada por "Para todo *a*, *R*, *b*, si él afirma que *a* tiene *R* para *b*, luego *a* tiene *R* para *b*". Resulta claro que si se incluyen "todas las formas de proposiciones" el análisis debe ser de carácter recursivo, pues las formas de las proposiciones siguen las formas (lógicas) de las oraciones, y hay un número infinito de éstas. No hay razón para suponer, entonces, que el análisis de Ramsey pueda completarse en una forma que no sea esencialmente paralela al método de Tarski para definir la verdad. El método de Tarski, sin embargo, introduce (me atrevo a afirmarlo) algo similar a la noción de correspondencia, y esto es precisamente lo que se suponía que las teorías que hemos estado explorando debían evitar. El proyecto recursivo de Ramsey puede verse también en problemas con la paradoja. Donde una teoría basada en el principio de (4) puede siempre pedir informalmente excusas porque un término de la forma "el enunciado de que *p*" sea incapaz de indicar cuándo una oración problemática reemplaza a "*p*", una teoría que pase revista sistemáticamente a las oraciones de un lenguaje necesitará apelar a un recurso más mecánico para evitar la contradicción. Uno se pregunta qué convicción trasuntaría la pretensión de Ramsey de que "en realidad no hay un problema de verdad aparte" después de que su análisis fuera llevado hasta su conclusión.

No he dicho nada acerca de los propósitos que se persiguen en una conversación (no filosófica) cuando se expresan oraciones que contienen "verdadero" y sus derivadas. Sin duda, la idea de que los enunciados que contienen la palabra "verdadero" son usados típicamente para expresar acuerdo, para enfatizar convicción o autoridad, para evitar repetición o para desviar responsabilidades, sumaría muchos partidarios si pudiera ponerse en evidencia que las palabras veritativas siempre pueden eliminarse sin merma cognitiva mediante la aplicación de una simple fórmula. No obstante, sostengo que las teorías acerca de los fines extralingüísticos de las oraciones que se expresan son lógicamente independientes de la cuestión de su significado; y es esto último lo que a mí me preocupa.

No hemos hallado una teoría satisfactoria que respalde la tesis de que es redundante atribuir verdad a los enunciados; pero aun si se pudiera demostrar que una teoría tal no es posible, esto no sería suficiente para establecer la teoría de correspondencia. Consideremos entonces más directamente las perspectivas de una explicación de la verdad en términos de correspondencia.

Lo que hace verdaderos los enunciados es la correspondencia entre lo que se dice y los hechos. Es natural, entonces, orientarse hacia los hechos en busca de ayuda. No puede aprenderse mucho de oraciones como:

(5) El enunciado de que Thika está en Kenia corresponde a los hechos

o de variantes del tipo de "Es un hecho que Thika está en Kenia", "Que Thika está en Kenia es un hecho", y "Thika está en Kenia, y eso es un hecho". Aceptemos o no la idea de que la correspondencia con los hechos explica la verdad, (5) y sus variantes no dicen más que "El enunciado de que Thika está en Kenia es verdadero" (o "Es verdadero que..." o "..., y eso es la verdad", etcétera). Si (5) llega a adquirir un interés independiente, será porque somos capaces de dar una explicación de los hechos y de la correspondencia que no lleva inmediatamente a la verdad. Una explicación así nos permitiría dar sentido a oraciones con esta forma:

(6) El enunciado de que *p* corresponde al hecho de que *q*.

El paso a la verdad sería simple: un enunciado es verdadero si hay un hecho al cual él corresponde. [(5) podría reescribirse "El enunciado de que Thika está en Kenia corresponde a un hecho"]. ¿Cuándo es válida (6)? Ciertamente cuando "*p*" y "*q*" son reemplazadas por la misma oración; más allá se enquistan las dificultades. El enunciado de que Nápoles está mucho más al norte que Red Bluff corresponde al hecho de que Nápoles está mucho más al norte que Red Bluff, pero también, se podría decir, al hecho de que Red Bluff está mucho más al sur que Nápoles (quizás ambos sean el mismo hecho). También corresponde al hecho de que Red Bluff está mucho más al sur que la más grande ciudad italiana en un radio de cincuenta kilómetros de Ischia. Cuando reflexionamos que Nápoles es la ciudad que satisface la siguiente descripción: es la ciudad más grande en un radio de cincuenta kilómetros de Ischia, y tal que Londres está en Inglaterra, entonces comenzamos a sospechar que si un enunciado corresponde a un hecho, corresponde a todos. ("Corresponde a los *hechos*" sería lo correcto en definitiva.) Por cierto, es fácil confirmar la sospecha empleando principios implícitos en nuestros ejemplos. Los principios son éstos: si un enunciado corresponde al hecho descrito por una expresión de la forma "el hecho de que *p*", luego corresponde al hecho descrito por "el hecho de que *q*" en tanto (1) las oraciones que reemplazan a "*p*" y a "*q*" sean lógicamente equivalentes, o (2) "*p*" difiera de "*q*" sólo en que el término singular ha sido reemplazado por un término singular coextensivo. El argumento confirmatorio es el siguiente. Supongamos que "*s*" abrevia a alguna oración verdadera. Luego, seguramente el enunciado de que *s* corresponde al hecho de que *s*. Pero podemos sustituir a la segunda "*s*" por la lógicamente equivalente "(*x* tal que *x* es idéntico a Diógenes y *s*) es idéntico a (*x* tal que *x* es idéntico a Diógenes)". Aplicando el principio según el cual podemos sustituir términos singulares coextensivos, podemos sustituir "*t*" por "*s*" en la última oración consignada, siempre que "*t*" sea verdadera. Finalmente, revirtiendo el primer paso concluimos que el enunciado de que *s* corresponde al hecho de que *t*, donde "*s*" y "*t*" son oraciones verdaderas cualesquiera.[6]

[6] Véase la nota 4 más atrás, y la nota 3 del Ensayo 2. Para una mayor discusión del argumento, y algunas sorprendentes aplicaciones, véase J. Wallace, "Propositional Attitudes and Identity".

Dado que aparte de cuestiones de correspondencia no se ha propuesto forma alguna de distinguir hechos, y dado que este test no logra descubrir una sola diferencia, podemos interpretar que el resultado de nuestros argumentos muestra que hay exactamente un hecho. Descripciones como "el hecho de que hay stupas en Nepal", si es que describen, describen la misma cosa: El Gran Hecho. No quedan razones para distinguir entre diversos nombres del Gran Hecho cuando se los escribe después de "corresponde a"; bien podemos recurrir a la frase única "corresponde a Gran Hecho". Este predicado inalterable arrastra consigo un redundante aire a ontología, pero aparte de esto parece no haber manera de distinguirlo de "es verdadero".

Una forma de disipar el argumento que llevó a esta conclusión sería rechazar los principios sobre los cuales se basa. E indudablemente uno puede imaginar el construir hechos de manera que podrían reflejar nuestra percepción del problema sin llevar a un colapso ontológico. Desde el punto de vista de la teoría de la verdad, sin embargo, todas esas construcciones parecen condenadas por la siguiente dificultad. Supongamos (reservando la sartén de la extensionalidad para los fuegos de la intensión) que podemos distinguir los hechos tan claramente como lo hacemos con los enunciados. Desde luego, no todo enunciado tiene su hecho; sólo los verdaderos lo tienen. Pero entonces, a menos que descubramos otra forma de distinguir los hechos, no podemos aspirar a explicar la verdad recurriendo a ellos.[7]

En los contextos que hemos considerado, la discusión acerca de hechos se reduce a una predicación de verdad; esto podría llamarse *teoría de redundancia* de los hechos. Las predicciones de verdad, por otra parte, no se han mostrado tan fáciles de eliminar. Si en esto no hay cabida para teorías de redundancia de la verdad, tampoco hay un estímulo para teorías de correspondencia.

Creo que nuestra frustración tiene una explicación bastante simple: hasta ahora no hemos tomado en cuenta el lenguaje. Los enunciados son verdaderos o falsos en razón de las palabras que se emplean en su construcción, y son las palabras las que tienen conexiones interesantes, detalladas, convencionales con el mundo. Cualquier teoría de la verdad seria debe entonces ocuparse de estas conexiones, y es precisamente aquí donde la noción de correspondencia puede tener algún predicamento. Nos hemos estado autorrestringiendo a formas de específicar enunciados que no hacen mención aparente de palabras. Así, "el enunciado de Jones de que el gato está sobre la alfombra" borra irremediablemente toda referencia a las particularidades del lenguaje de Jones que podrían conllevar una consideración no trivial de la verdad, y puede pensarse que en general esto sucede con la frase "el enunciado de que *p*".

Las discusiones sobre la verdad pueden haber evitado el aspecto lingüístico debido a que resulta obvio que la verdad no puede depender

[7] Hay una observación similar en P. F. Strawson, "Truth: A Reconsideration of Austin's Views".

de creaciones; pero si éste ha sido un motivo, es un motivo confuso. Las oraciones no pueden ser verdaderas o falsas pues si lo fueran tendríamos que decir que "Je suis Titania" es verdadera (si la dice o la canta Titania), falsa (si la dice cualquier otra persona), y ninguna de ambas (si la emite alguien que no habla francés). Lo que este familiar argumento revela no es que debemos dejar de hablar simultáneamente de oraciones y de verdad, si no que debemos profundizar un poco más y hablar también del momento (tiempo) en que se emite la oración, y de su emisor.

La verdad (en un lenguaje natural dado) no es una propiedad de las oraciones. Es una relacion entre oraciones, hablantes y fechas. Verlo de esta manera no significa apartarse del lenguaje hacia mudas entidades eternas como proposiciones, enunciados y aserciones, sino relacionar el lenguaje con las ocasiones de la verdad en una forma que conduzca a la construcción de una teoría.

Los dos últimos párrafos pueden sugerir que para tener una teoría competente acerca de la verdad debemos abandonar la idea de que los enunciados son los vehículos adecuados de la verdad. Pero este no es el caso. Si yo estoy en lo cierto, las teorías de la verdad deben caracterizar o definir un predicado triádico "T *s, u, t*". No interesará a la teoría que leamos este predicado como "la oración *s* es verdadera (en castellano) para el hablante *u* en el tiempo *t*" o como "el enunciado expresado por la oración *s* (en castellano) por el hablante *u* en *t* es verdadero". Quienes creen que, por otras razones, debemos conservar los enunciados como vehículos de la verdad hallarán más perspicua la segunda formulación, con su término singular complejo ("El enunciado...") y su predicado monádico ("es verdadero"), mientras quienes (como yo) creen que podemos habérnosla sin enunciados pueden preferir la primera formulación, más austera. Pero ambas partes puede emplear ambas formas; la diferencia sólo aparece cuando la discusión se ve a la luz de una teoría amplia. Las cuestiones en discusión no establecen, pienso, si dicha teoría requiere o no una ontología de enunciados.

Hay excelentes razones para no predicar la verdad de oraciones, pero estas razones no son aplicables a actos de habla, emisiones o instancias particulares. Se ha argumentado, y en forma convincente, que casi nunca o quizá nunca decimos de un acto de habla, de una emisión o de una instancia particular, que es verdadero.[8] Esto apenas señala por qué no *debemos* llamar verdaderas a estas entidades (si es que existen). No se generaría confusión si dijéramos que el particular habla de una oración fue verdadero en caso de haber sido usada en tal ocasión para hacer un enunciado verdadero; esto sería similar para instancias particulares y para emisiones. De acuerdo con Strawson:

> "Mi enunciado" puede ser tanto lo que digo como mi decir eso. Mi decir algo es ciertamente un episodio. Lo que digo no lo es. Es esto último, y no lo anterior, lo que declaramos ser verdadero.[9]

[8] Véase R. Cartwright, "Propositions".
[9] P. F. Strawson, "Truth", 129-30.

No estoy seguro de que un enunciado sea siempre un acto de habla, pero en todo caso podemos aceptar la conclusión de que no se dice de los actos de habla que sean verdaderos. Pero ¿qué resulta de esto? Sin duda, no resulta la imposibilidad de explicar qué es hacer un enunciado verdadero en términos de las relaciones convencionales entre palabras y cosas que valen para cuando las palabras son usadas por agentes particulares en ocasiones particulares. Pues si bien "mi enunciado" puede no referir, al menos cuando se cuestiona la verdad, a un acto de habla, aún puede tener éxito en la identificación de su enunciado con sólo relacionarlo con un acto de habla. (¿Qué lo convierte en "mi" enunciado?)

Si alguien que habla castellano emite la oración "El sol está sobre el palo mayor", ¿bajo qué condiciones ha hecho el enunciado de que el sol está sobre el palo mayor? Un rango de respuestas podría incluir estipulaciones tales como que intentaba transmitir a sus oyentes la impresión de que él creía que el sol estaba sobre el palo mayor, que su status lo autorizaba a producir información acerca de la posición del sol, etcétera. Siguiendo esa línea de pensamiento, uno podría sostener que, si el hablante no tenía idea de la posición del sol, y quería anunciar que era tiempo de un trago, luego *no* hizo el enunciado de que el sol está sobre el palo mayor. Pero también en cierto sentido tendríamos que decir que se hizo un enunciado, incluso bajo las condiciones que acabamos de mencionar, por cuanto el hombre había hecho ("literalmente") el enunciado de que el sol estaba sobre el palo mayor, y lo que él dijo era ("literalmente") verdadero puesto que el sol estaba sobre el palo mayor en el momento en que él hablaba, aun cuando él no tenía razones para creerlo, y no le importaba que fuera verdadero. En estos casos, no nos interesa lo que la persona quiso significar al emitir la oración, sino lo que la oración, tal como fue emitida, significaba. Ambas nociones de significado son relativas a las circunstancias de la ejecución, pero en el segundo caso hacemos abstracción de las intenciones extralingüísticas del hablante. La comunicación por medio del lenguaje es la comunicación por medio del significado literal; así debe haber un sentido literal de hacer un enunciado, si es que hay otros. La teoría de la verdad se ocupa del sentido literal. (Desde luego, este tema merece una discusión mucho más extensa.)

Ajustándonos a lo literal, entonces, alguien que habla castellano hará un enunciado verdadero al emitir la oración "Es martes" si y sólo si en el momento en que habla a su alrededor es martes. Este ejemplo invita a la generalización: cada instancia del siguiente esquema será una verdad acerca de la verdad cuando se reemplaza "*s*" por una descripción de una oración del castellano y se reemplaza "*p*" por una oración que da las condiciones bajo las cuales la oración descrita es verdadera:

(7) La creación *s* es verdadera (en castellano) para el hablante *u* en el tiempo *t* si y sólo si *p*.[10]

[10] La expresión que reemplaza a "*p*" debe contener a "*u*" y a "*t*" como variables libres a menos que no haya elementos deícticos en *s*.

(Podría sustituirse por un esquema alternativo que aparentemente atribuye verdad a los enunciados.) Aun si restringimos las descripciones que sustituyen "*s*" a algún estilizado vocabulario de sintaxis, podemos considerar que hay una oración verdadera de la forma de (7) para cada oración castellana. La totalidad de tales oraciones determina la extensión del predicado triádico de (7) (el predicado de verdad relativizado). Parecemos estar aquí al borde de una teoría de la verdad; pero todavía no se divisa nada parecido a la correspondencia. La razón de esto puede ser, sin embargo, que estamos *sólo* al borde de una teoría. El esquema (7) nos dice qué debe tener una teoría de la verdad, pero no es en sí mismo tal teoría, y no sugiere cómo puede construirse esa teoría. El esquema (7) está concebido para desempeñar en el castellano un papel análogo al que desempeña en un lenguaje artificial un esquema similar a la Convención T de Tarski.[11] Ese papel consiste en proporcionar un test de la adecuación de una teoría de la verdad: una teoría aceptable debe implicar una oración verdadera de la forma de (7) independientemente de cuál sea la oración del castellano descrita por la expresión canónica que reemplaza a "*s*".

El esquema (7) carece de una elegante característica de su análogo en la obra de Tarski. Este, al no interesarse por los lenguajes con elementos deícticos, puede usar esta simple fórmula: "*s* es verdadero (en *L*) si y sólo si "*p*" donde la oración que sustituye a "*p*" es la oración descrita por la expresión que reemplaza a "*s*" si el metalenguaje contiene al lenguaje objeto; de lo contrario éste traduce íntegramente esa oración. Esta sencilla fórmula no puede sernos útil, pues cuando hay términos deícticos (demostrativos, tiempos verbales) lo que ocupa el lugar de "*p*" en general no puede ser lo que "*s*" nombra o una traducción de ello, tal como lo atestigua el ejemplo de la primera oración del párrafo precedente. La elaboración que se requeriría para expresar (7) en términos sintácticos explícitos sería considerable, pero no hay razones para considerarlo imposible dado que la reemplazante de "*p*" debe ser sistemáticamente relacionada con la oración descrita por la reemplazante de "*s*" mediante las reglas que gobiernan el uso de los términos deícticos en castellano.

Si las oraciones indicativas del castellano comprendieran tan sólo un numero finito de oraciones elementales y de compuestos veritativos funcionales de ellas, sería fácil efectuar una caracterización recursiva de la verdad: proveyendo una oración de la forma de (7) para cada oración elemental y dando una regla que corresponda a cada conectivo oracional. Sin embargo, esta estrategia se derrumba tan pronto como permitimos la conformación de predicados de complejidad arbitraria mediante el uso de variables y conectivos, como en la cuantificación o en los términos singulares complejos; y es precisamente aquí donde la teoría de la verdad se torna interesante. Concentrémonos en la estructura cuantificacional a expensas de los términos singulares, no sólo por cau-

[11] A. Tarski, "The Concept of Truth in Formalized Languages", 187, 188.

sa de que los últimos pueden considerarse prescindibles mientras que la primera no, sino también porque la cuestión que surgirá puede resolverse con mayor sencillez. El problema que presenta la estructura cuantificacional para una teoría recursiva de la verdad es, desde luego, que si bien pueden construirse oraciones de cualquier longitud finita a partir de una pequeña dotación de variables, conectivos, predicados y cuantificadores, ninguna de las partes de una oración necesita ser a su vez una oración; de allí que la verdad de una oración compleja en general no pueda inferirse en términos de la verdad de sus partes.

Tarski nos enseñó a percibir el problema y nos dio una ingeniosa solución. La solución depende de caracterizar primero una relación llamada *satisfacción* y de definir luego la verdad mediante ella. Las entidades que se satisfacen son oraciones tanto abiertas como cerradas; los satisfactores son funciones que relacionan las variables del lenguaje objeto con las entidades que abarcan: casi todo, si el lenguaje es el inglés.[12] Una función satisface a un predicado inestructurado n-ádico con variables en sus n espacios si el predicado es *verdadero para* las entidades (en orden) que la función asigna a esas variables. Así, si "x ama a y" es una oración abierta del tipo más simple, una función f la satisface sólo en caso de que la entidad que f asigna a "x" ame a la entidad que f asigna a "y". A su vez, la caracterización recursiva de la satisfacción debe pasar revista a todos los predicados primitivos. Esta hace frente a los conectivos de la manera obvia: así, una conjunción de dos oraciones s y t (abiertas o cerradas) es satisfecha por f siempre que f satisfaga a s y f satisfaga a t. La cuantificación universal de una oración abierta s con respecto a una variable v es satisfecha por f en caso de que f, y toda otra función como f a excepción de aquello que ésta asigne a v, satisfaga a s. (Esta oración también funciona si "existencial" reemplaza a "universal" y si "o alguna" reemplaza a "y toda".) Que una función particular satisfaga o no a una oración depende completamente de cuáles sean las entidades que aquélla asigne a las variables libres de la oración. Luego si la oración no tiene variables libres —si es una oración cerrada, o genuina— entonces debe ser satisfecha por todas las funciones o por ninguna. Y, como según los detalles de la recursión resulta claro, aquellas oraciones cerradas que son satisfechas por todas las funciones son verdaderas; aquellas que no son satisfechas por ninguna son falsas. [En todo momento entiendo que la satisfacción, al igual que la verdad, está relativizada al estilo de (7).]

El concepto semántico de la verdad desarrollado por Tarski merece ser considerado como una teoría de correspondencia a causa de la función que desempeña el concepto de satisfacción; pues se ha podido explicar así, y de manera no trivial, la propiedad de ser verdadero en términos de una relación entre lenguaje y algo más. La relación, que lla-

[12] Los satisfactores de Tarski son secuencias infinitas, no functores. A un lector en busca de precisión y de una comprensión más profunda no debe insistírsele mucho para que estudie "The Concept of Truth in Formalized Languages".

mamos satisfacción, no es, admitámoslo, exactamente lo que la intuición esperaba de la correspondencia; y las funciones o secuencias que se satisfacen pueden no parecerse mucho a hechos. En parte el contraste se debe a una característica especial de las variables: justamente a causa de que no refieren a ningún individuo en particular, la satisfacción debe considerar asignaciones arbitrarias de entidades a las variables (nuestras funciones). Si en cambio pensáramos en nombres propios, los satisfactores podrían acercarse más a los objetos ordinarios de nuestro discurso: es decir, serían *n*-tuplos ordenados de ellos. Así "Dolores ama a Dagmar" sería satisfecha por Dolores y por Dagmar (en ese orden), siempre que Dolores amara a Dagmar. Supongo que Dolores y Dagmar (en ese orden) no es también un hecho; el hecho que verifica a "Dolores ama a Dagmar" debe incluir de alguna forma el amor. Este "de alguna forma" ha sido siempre la némesis de las teorías de la verdad basadas en los hechos. Entonces el punto en cuestión no es que "*s* es satisfecha por todas las funciones" signifique exactamente lo mismo que pensábamos que significaba "*s* corresponde a los hechos", sino que las dos frases tienen esto en común: ambas procuran expresar una relación entre el lenguaje y el mundo, y ambas son equivalentes a "*s* es verdadera" cuando *s* es una oración (cerrada).

La comparación entre teorías de correspondencia que explotan el concepto de satisfacción y aquellas que se apoyan en la correspondencia con los hechos tiene su pico máximo en las oraciones sin variables libres. El paralelo aún se extiende —si aceptamos el argumento de Frege acerca de la extensión de las oraciones— a la conclusión de que las oraciones verdaderas no pueden ser diferenciadas en base a aquello a que ellas corresponden (los hechos, El Gran Hecho) o aquello que las satisface (todas las funciones, secuencias). Pero la estrategia de Tarski puede afrontar este parecido en el producto terminado donde la estrategia de los hechos no puede hacerlo, puesto que la satisfacción de oraciones cerradas se explica en términos de satisfacción de oraciones tanto abiertas como cerradas, en tanto que sólo las oraciones cerradas tienen tradicionalmente hechos correspondientes. Dado que diferentes asignaciones de entidades a las variables satisfacen diferentes oraciones abiertas y dado que las oraciones cerradas se construyen a partir de las abiertas, en la aproximación semántica la verdad se alcanza por diferentes caminos para diferentes oraciones. Todas las oraciones verdaderas terminan en el mismo lugar, pero las historias de su arribo hasta allí son diferentes; una teoría semántica de la verdad cuenta la historia para una oración particular recorriendo los pasos de la consideración recursiva de la satisfacción apropiada para la oración. Y la historia constituye una prueba de un teorema bajo la forma de una instancia de esquema (7).

La estrategia de los hechos no puede proporcionar tanta variedad instructiva. Dado que todas las oraciones verdaderas tienen la misma relación con los hechos, una explicación de la verdad de una oración en base a sus relaciones con otras oraciones (cerradas) debe, si se atiene a los hechos, comenzar por donde termina.

Visto en retrospectiva, el fracaso de las teorías de correspondencia

de la verdad basadas en la noción de hecho puede remontarse a una fuente común: el deseo de incluir en la entidad a la cual corresponde una oración verdadera no sólo los objetos "acerca de los cuales" es la oración (otra idea llena de problemas) sino también toda cosa que la oración diga acerca de ellos. Una consecuencia bien explorada de esto es que se hace difícil describir el hecho que verifica una oración excepto mediante el uso de la propia oración. La otra consecuencia es que la relación de correspondencia (o "ilustración") sólo parece tener aplicación directa para las oraciones más simples ("Dolores ama a Dagmar"). Esto impulsa a los teóricos del hecho a tratar de explicar la verdad de todas las oraciones en términos de la verdad de las más simples y de aquí, en particular, a interpretar la cuantificación como una mera taquigrafía de conjunciones o alternaciones (quizá de longitud infinita) de las oraciones más simples. La ironía es que, en tanto veamos a la cuantificación bajo esta luz, no hay una necesidad real de algo como la correspondencia. Sólo cuando nos vemos forzados a tomar a la generalidad como un agregado esencial para los recursos conceptuales de la predicación y para la composición de oraciones, y como algo no reducible a ellos, podemos apreciar los usos que tiene una teoría de correspondencia refinada. Una teoría de la verdad basada en la satisfacción resulta instructiva en parte porque es menos ambiciosa respecto de las atribuciones que da a las entidades a las cuales corresponden las oraciones: en una teoría de este tipo las entidades no son más que pares arbitrarios formados por los objetos abarcados por las variables del lenguaje y por esas variables. La relativa simplicidad de los objetos se ve compensada por el trabajo que insume explicar la relación entre ellos y las oraciones, pues toda la característica veritativa relevante a toda oración debe ser tenida en cuenta al describir la satisfacción. El beneficio es claro: al explicar la verdad en términos de satisfacción, todos los recursos conceptuales del lenguaje en relación a su ontología son inducidos a fructificar.

Hablar de oraciones, o mejor dicho enunciados, fieles a, o correspondientes a los hechos, es sin duda tan inofensivo como hablar de la verdad. Incluso la sugerencia que encierran estas frases de que la verdad se debe a una relación entre el lenguaje y el mundo puede llegar a justificarse, tal como ya lo he sostenido. La estrategia de los hechos, a la que he venido atacando, es algo diferente: una teoría filosófica, y una de las malas. Sería una vergüenza desacreditar a todas las teorías de correspondencia, y en particular a la aproximación semántica de Tarski, adjudicándoles las mismas imperfecciones que pululan en los intentos habituales de explicar la verdad en base a los hechos.

La presunción de que todas las teorías de correspondencia deben hacer uso de la estrategia de los hechos es al menos comprensible y, dadas las extravagancias de la práctica filosófica, podría considerarse verdadera por decreto. Es menos excusable la difundida incomprensión del papel de fórmulas como (7) en la aproximación semántica. El siguiente ejemplo no es peor que muchos otros que podrían citarse:

...a menos que la "correspondencia" en la que tanto insisten las teorías de correspondencia de la verdad clásicas encierre algo más que lo que pueden captar las formulaciones de la actual teoría semántica y a menos que este algo más pueda dar evidencia de ser una propiedad esencial de la verdad (o, al menos, de una variedad significativa de verdades), la batalla por la correspondencia, en vez de dar el *triunfo* a los teóricos de la correspondencia, ha probado ser un *Scheinstreit*. Pues, como se ha notado a menudo, la fórmula

"La nieve es blanca" (en nuestro lenguaje) es verdadero = La nieve es blanca,

es considerada con la mayor ecuanimidad tanto por el pragmatista como por el coherentista. Si la "correspondencia" del teórico de la correspondencia equivale nada más que a lo que ilustran tales equivalencias, entonces, mientras el pragmatista y el coherentista pretenderán hacer importantes aportes... nada más quedará por decir acerca de "la verdad y la correspondencia".[13]

Tenga o no la aproximación semántica a la verdad algo más que lo que Sellars está dispuesto a admitir, quizá sea el caso que ninguna batalla se gane, o siquiera se libre, entre las teorías de correspondencia y las demás. Mi preocupación por este pasaje surge de su presunción de que una oración como " 'La nieve es blanca' es verdadera si y sólo si la nieve es blanca" (inclusive después de ser convenientemente relativizada y con una descripción estructural en el lugar de la cita) proporciona por sí misma una pista de lo que es privativo de la aproximación semántica. Desde luego, como dice Sellars, tales oraciones son terreno neutral; es precisamente por esta razón que Tarski tiene esperanzas de que todos puedan coincidir en que una teoría o definición adecuada de la verdad debe implicar a todas las oraciones de esta forma. No hay trazas de la noción de correspondencia en estas oraciones, no hay un predicado relacional que exprese una relación entre las oraciones y aquello de que ellas tratan. Donde esa relación, la satisfacción, *sí* entra en juego es en la elaboración de una teoría no trivial capaz de afrontar la prueba de implicar todas esas trivialidades neutrales cubiertas de nieve.

Quisiera ahora, a modo de conclusión, mencionar brevemente dos de los muchos tipos de obstáculos que habrá que superar si es que vamos a tener alguna vez una amplia teoría de la verdad para un lenguaje natural. Primero, es por cierto razonable imaginar hasta qué punto será alguna vez posible tratar un lenguaje natural como un sistema formal, y mucho más lo es cuestionarse si los recursos del método semántico podrán comenzar a abarcar fenómenos comunes tales como la modificación adverbial, los adjetivos atributivos, el discurso de las actitudes proposicionales, de la causalidad, de la obligación, y todos los demás. Al presente ni siquiera tenemos una semántica satisfactoria para los términos

[13] W. Sellars, "Truth and 'Correspondence' ", 29.

singulares. No obstante, cierto grado de optimismo es justificable. Hasta Frege, la semántica seria se limitaba mayormente a la predicación y a la composición veritativa funcional de oraciones. Al abstraer la estructura cuantificacional de lo que parecía ser una jungla de pronombres, cuantificadores, conectivos y artículos, Frege mostró de qué manera podía dominarse semánticamente un fragmento asombrosamente poderoso del lenguaje natural. Por cierto, todavía puede resultar que este fragmento demuestre, ingenio mediante, ser el todo. Mientras tanto, prometedores trabajos avanzan en muchas direcciones, incrementando los recursos de la semántica formal, extendiendo la aplicación de recursos conocidos, y proveyendo las complejas y detalladas reglas necesarias para dar una descripción reveladora de la estructura del lenguaje natural. Cualquiera que sea el alcance que finalmente demuestre tener la teoría semántica de la verdad, la percepción que resulta de conocer el lenguaje lo suficientemente bien como para poder aplicarla merece ser celebrada.

La segunda dificultad pertenece a otro nivel: hemos sugerido cómo sería posible interpretar las atribuciones de verdad a enunciados o a oraciones relativizadas a ocasiones de uso, pero sólo en contextos de tipo presentado por la rama izquierda de (7). No hemos indicado de qué manera podría extenderse el análisis para aplicarse a oraciones como:

(8) Es verdadero *that* está lloviendo.
(9) El enunciado *that* está lloviendo es verdadero.

Aquí vemos cómo podríamos tratar de hacer frente al caso de (8). Tenemos, supóngase, una teoría de la verdad para el inglés que trata a la verdad como una relación entre una oración, un hablante y un tiempo. [La versión alternativa en términos de enunciados se aplicaría a (9).] El problema es hallar contrapartes naturales de estos elementos de (8). Un hablante de (8) dice las palabras "está lloviendo", cumpliendo así un acto que involucra una oración particular, tiene su hablante, y su tiempo. Una referencia a este acto puede así servir como referencia a los tres ítems que se necesitan para aplicar la teoría de la verdad. Podemos pensar que la referencia ha sido resumida al demostrativo "*that*" de (8) y (9). Podríamos devanar a (8) del siguiente modo. Primero (invirtiendo el orden para mayor claridad), digo: "Está lloviendo". Luego digo "*that* acto de habla incluía una oración que, dicha por mí ahora, es verdadera". En este análisis, una emisión de (8) o (9) consiste en dos actos de habla lógicamente (semánticamente) independientes, uno de los cuales contiene una referencia demostrativa del otro. Rasgo característico de estas emisiones es que una es verdadera si la otra lo es; quizás esto confirma una idea de la teoría de redundancia.

Un problema adicional surge de

(10). El enunciado de Pedro *that* Pablo es rubio es verdadero.

Siguiendo la sugerencia hecha para (8) y (9), el análisis de (10) debería ser "Pablo es rubio. *That* es verdadero, y Pedro lo dijo (lo enunció)". El "*that*" refiere, como antes, a un acto de habla, y ahora él "lo" recoge la

misma referencia. Para completar la consideración hace falta un análisis paratáctico del discurso indirecto que interprete que una emisión por un hablante *u* de "Pedro dijo *that* Pablo es rubio" está compuesta por una emisión de "Pablo es rubio" y otra emisión ("Pedro dijo *that*") que relaciona a Pedro en una cierta forma con la emisión de *u* "Pablo es rubio". La relación en cuestión tal vez pueda hacerse inteligible si apelamos a la noción de *igual-decir*: si *u* dice lo que es verdadero cuando dice "Pedro dijo *that*", esto es porque, al decir "Pablo es rubio", ha convertido a Pedro y a sí mismo en igual-dicentes.[14]

Por supuesto, uno puede insistir en que la relación de igual-decir (que es aplicable entre actos de habla) sólo puede entenderse en referencia a una tercera entidad: un enunciado, un significado o una proposición. Nada de lo que aquí he escrito se relaciona con esta cuestión, salvo indirectamente al mostrar que, con respecto a los problemas en discusión, no surge una necesidad de tales entidades. ¿Es esto simplemente el resultado de dejar de lado los casos problemáticos? Consideraremos, como ejemplo final:

(11) Pedro dijo algo verdadero.

Esto no puede transformarse en "Alguna emisión (pasada) de Pedro nos convierte en igual-dicentes", pues yo puedo no haber dicho, o no saber cómo decir, la cosa apropiada. Tampoco ayudará probar con "Alguna emisión de Pedro incluyó una oración verdadera bajo las circunstancias". Esto fracasa pues (11) no dice en qué lenguaje habló Pedro, y el concepto de verdad con que estamos operando está necesariamente limitado a un lenguaje específico. Al no saber qué lenguaje es, no podemos dar un sentido de "verdadero en su lenguaje".

Sí podemos confiar en interpretar, creo, la idea de una oración en otra lengua como una *traducción* de una oración del inglés. Dada esta idea, resulta natural ver que (11) significa algo como "Pedro emitió una oración que traduce una oración de inglés verdadera bajo las circunstancias". La naturaleza exacta de la suposición contrafáctica apenas disimulada en este análisis depende de los detalles de la teoría de la verdad (para el inglés) relativizada para las ocasiones de emisión. En todo caso, parece que se nos exige comprender lo que otra persona querría significar con una oración de nuestro lenguaje si hablara nuestro lenguaje. Pero por embarazoso que sea este concepto, es difícil ver cómo puede existir la comunicación sin él.

La conclusión que yo tentativamente extraería es la siguiente. Podemos evadirnos de lo que aparentemente sería hablar de la verdad (absoluta) de enunciados atemporales si aceptamos una verdad relativizada a las ocasiones de habla y una poderosa noción de traducción. Esta opción puede crear más problemas que los que resuelve. Pero ellos son, creo yo, los problemas correctos: proporcionar una descripción detallada

[14] Para más detalles sobre esta aproximación al discurso indirecto, véase el Ensayo 7.

de la semántica del lenguaje natural, e idear una teoría de la traducción que no dependa de, sino que más bien fundamente, lo que quiera que sea el concepto de significado.

Strawson describe la "versión purificada de la teoría de correspondencia de la verdad" de Austin en los siguientes términos:

> Su... teoría es, a grandes rasgos, que decir que un enunciado es verdadero es decir que cierto episodio de habla está relacionado de cierta manera convencional a algo en el mundo exclusivo de sí mismo.[15]

Es ésta la teoría que Strawson tiene en mente cuando dice: "La teoría de correspondencia no requiere purificación, sino eliminación". No quisiera defender los pormenores de la concepción de correspondencia de Austin, y muchos de los puntos que he enumerado en contra de la estrategia de los hechos reflejan las críticas de Strawson. Pero las debilidades de formulaciones particulares de la teoría de correspondencia no deben arrojarse contra la teoría. Si yo estoy en lo cierto, apelando a la concepción semántica de la verdad de Tarski podemos defender una teoría que coincide casi exactamente con la descripción que Strawson hace de la "versión purificada de la teoría de correspondencia de la verdad" de Austin. Y esta teoría merece no su eliminación, sino su elaboración.

[15] P. F. Strawson, "Truth", 129.

4

Semántica para lenguajes naturales

Una teoría de la semántica de un lenguaje natural tiene por objeto dar el significado de toda expresión significativa, pero está en cuestión la forma que una teoría debería tomar para lograrlo. Puesto que el número de expresiones significativas parece no tener un límite claro, una teoría posible debe dar cuenta del significado de cada expresión sobre la base de la exhibición pautada de un número finito de rasgos. Aunque haya un límite práctico para la longitud de las oraciones que una persona puede comprender al emitir y recibir, una semántica satisfactoria necesita explicar la contribución de los rasgos repetibles al significado de las oraciones en las que ellos aparecen.

Creo que una teoría de la verdad para un lenguaje sí hace, de un modo mínimo pero importante, lo que nosotros queremos, esto es proporciona los significados de todas las expresiones independientemente significativas en base a un análisis de su estructura. Y por otra parte, una teoría semántica de un lenguaje natural no puede considerarse adecuada a menos que proporcione una descripción del concepto de verdad para ese lenguaje según los lineamientos fundamentales propuestos por Starsky para los lenguajes formalizados. Pienso que tanto los lingüistas como los filósofos interesados en los lenguajes naturales han pasado por alto la importancia fundamental de la teoría de la verdad en parte porque o se han dado cuenta de que una teoría de la verdad da una respuesta precisa, profunda y comprobable a la cuestión acerca de la forma en que los recursos finitos resultan suficientes para explicar las infinitas capacidades semánticas del lenguaje, y en parte porque han exagerado las dificultades del proceso de construcción de una teoría formal de la verdad para un lenguaje natural. De todas maneras intentarlo resulta instructivo, pues observamos que en cuanto tengamos éxito en proveer una teoría tal a un lenguaje natural, veremos al lenguaje natural como un sistema formal; y en cuanto nos pongamos por meta la construcción de esa teoría, podremos ver a lingüistas y filósofos analíticos como colaboradores.[1]

Por teoría de la verdad entiendo un conjunto de axiomas que implica, para cada oración del lenguaje, un enunciado de las condiciones bajo las cuales ella es verdadera. Obviamente, si tenemos una definición de

[1] Véase la nota 14 del Ensayo 2.

un predicado de verdad que satisface la Convención T de Tarski, tenemos una teoría de la verdad [2], pero en general se requiere menos para la caracterización de una teoría de la verdad. Si no se imponen más restricciones, algunas teorías de la verdad resultarán de poco interés intrínseco. Por ejemplo, podríamos tomar simplemente por axiomas a todas las oraciones de la forma "*s* es verdadero si y sólo si *p*" donde "*s*" sea reemplazada por una descripción estadarizada de una oración y "*p*" por esa oración (suponiendo que el metalenguaje contenga al lenguaje objeto). Una teoría así no podría ofrecer una visión de la estructura del lenguaje y entonces tampoco daría un asomo de respuesta a la cuestión de la forma en que el significado de una oración depende de su composición. Podríamos eliminar esta aberración en particular si estipuláramos que los axiomas no lógicos deben ser finitos en número; en lo que sigue supondré que esta restricción ya rige, si bien podrían existir otras maneras de asegurar que una teoría de la verdad posee las propiedades que nosotros queremos. ¿Qué propiedades queremos? Como ya hemos dicho, una teoría aceptable debería dar cuenta del significado (o de las condiciones de verdad) de toda oración analizando su composición, según formas de la verdad relevante, a partir de elementos tomados de un repertorio finito.[3] Un segundo requisito natural es que la teoría proporcione un método para decidir, dada una oración arbitraria, cuál es su significado. (De una teoría que satisfaga estas dos condiciones puede decirse que demuestra que el lenguaje que describe es *aprendible* y *escrutable.*) Una tercera condición es que los enunciados de condiciones de verdad para oraciones individuales incluidas en la teoría deberían, de alguna forma todavía por precisar, hacer uso de los mismos conceptos que usan las oraciones cuyas condiciones de verdad estipulan.[4]

Parece claro que las teorías del tipo de Tarski mostró cómo idear gozan de estas características deseables. Por ejemplo, una teoría que descansa en un metalenguaje que contiene al lenguaje objeto satisface en una forma elemental la última condición, pues en los enunciados de la forma "*s* es verdadero si y sólo si *p*" que tratamos las condiciones de verdad de *s* están dadas por la oración que reemplaza a "*p*", es decir *s* misma, y así no hacen uso de ningún concepto que no haya sido requerido directamente para la comprensión de *s*. Si el metalenguaje no contiene al lenguaje objeto, la satisfacción de este criterio parece menos obvia; y los lenguajes naturales generan otros problemas que trataremos oportunamente.

Parece natural interpretar que la tercera condición prohíbe la apa-

[2] A. Tarski, "The Concept of Truth in Formalized Languages".

[3] Condiciones de verdad y significados no se considerarán equivalentes; a lo sumo podemos decir que al dar las condiciones de verdad de una oración estamos dando su significado. Pero esto también necesita clarificación. Para algunos requisitos necesarios, véanse los Ensayos 9 y 12. (Nota agregada en 1982.)

[4] Para importantes reservas en la posibilidad de precisar esta condición, véase S. Kripke, "Is There a Problem about Substitutional Quantification?" (Nota agregada en 1982.)

ricion de un término semántico en el enunciado de las condiciones de verdad de una oración a menos que la oración ya contenga al término semántico (o a una traducción de él). No queda claro si esto excluiría o no un recurso explícito a conceptos semánticos en el enunciado de las condiciones de verdad para oraciones modales (puesto que es incierta la determinación desde un principio de su naturaleza semántica). Pero esta limitación más bien parece amenazar a las teorías que apelan a un concepto no analizado de lo que es denotar o nombrar, así como a aquellas que convierten a la verdad en la noción semántica fundamental de un modelo.[5]

Pareciera ser que asignar esta interpretación a la tercera condición equivale a juzgar irrelevante para nuestros propósitos presentes a buena parte de los trabajos recientes en torno de la semántica; por ello pretendo dejar esta cuestión abierta, junto a muchas otras cuestiones tocantes a la formulación detallada de la pautas que deberíamos exigir de una teoría de la verdad. Mi interés actual no se centra en la discusión de puntos en disputa sino en destacar la importancia y la utilidad que en general redundaría de requerir a toda teoría del significado (semántico) que proporcione una descripción recursiva de la verdad. Creo que no exageraríamos los méritos de esta idea al decir que nos daría un marco dentro del cual podríamos determinar nítidamente una multitud de beneficios y problemas.

Dar una teoría recursiva de la verdad para un lenguaje es mostrar que la sintaxis del lenguaje es formalizable, al menos en el sentido de que al analizar todas las expresiones verdaderas se podrá determinar que están formadas por elementos (el "vocabulario"), una provisión finita de los cuales bastará para el lenguaje por la aplicación de reglas, un número finito de las cuales bastará para el lenguaje. Pero aun si llegamos a suponer que la falsedad puede definirse en términos de la verdad, o que puede caracterizarse de manera independiente o similar, no se desprende de ello que la oracionalidad o la gramaticalidad puedan definirse recursivamente. Así, los argumentos destinados a establecer que no puede darse una descripción formal recursiva de la sintaxis (oracionalidad o gramaticalidad) de un lenguaje natural no necesariamente desacreditan el intento de proporcionar una teoría de la verdad. También debería mencionarse que las condiciones de adecuación sugeridas para una teoría de la verdad no implican (y esto es obvio) que también la oraciones verdaderas del lenguaje objeto tengan la forma de algún sistema lógico habitual. Aun suponiendo que fuera claro (lo que no ocurre) que la estructura profunda del castellano (u otro lenguaje natural) no puede ser representada por un lenguaje formal con la estructura cuantificacional usual, tampoco se podría concluir que no hay manera de dar una teoría de la verdad.

Una teoría de verdad para un lenguaje natural debe tener en cuen-

[5] Véase J. Wallace, "Nonstandard Theories of Truth".

ta el hecho de que muchas oraciones varían su valor de verdad dependiendo del tiempo en que se las dice, del hablante e incluso, quizá, de la audiencia. Podemos dar cabida a este fenómeno sea declarando que son las emisiones particulares o los actos de habla, y no las oraciones, las que tienen valores de verdad, sea haciendo de la verdad una relación aplicable entre una oración, un hablante y un tiempo. Adecuar de este modo a los elementos deícticos, o demostrativos, en un lenguaje natural es aceptar un cambio conceptual radical en la forma en que puede definirse la verdad, como podrá apreciarse al reflejar la forma en que debe revisarse la Convención T para hacer a la verdad sensible al contexto. Pero el cambio no significa necesariamente apartarse de la formalidad.

A menudo se expresa el temor de que una teoría formal de la verdad no sea capaz de sobrellevar los problemas de ambigüedad del lenguaje natural que tanta energía insumen a los lingüistas. Cuando se piensa en esto puede ser de ayuda distinguir dos opiniones. Una es que las teorías de la verdad formales no han sido concebidas por lo general para hacer frente a la ambigüedad, y el tener que equiparlas para hacerlo alteraría su naturaleza. Esta opinión está justificada, y es inofensiva. En general, la teorías de la verdad al estilo de Tarski no tratan cuestiones de definición para el vocabulario primitivo (a diferencia de las cuestiones de traducción y de forma lógica); por otra parte, nada en una teoría de la verdad es enemigo del tratamiento satisfactorio de los problemas para cuya solución se ha creado un léxico. La segunda opinión dice que algunos tipos de ambigüedad impiden necesariamente que proporcionemos una teoría de la verdad. Antes de que podamos evaluar esta tesis será necesario ser mucho más claros que hasta ahora acerca de los criterios de éxito respecto de dar una teoría de la verdad para un lenguaje natural. Sin pretender aquí una discusión profunda, permítaseme señalar por qué pienso que este asunto no puede resolverse citando unos cuántos casos problemáticos.

Bar-Hillel da un ejemplo como éste: "Vinieron en lentos trenes y aviones".[6] Podemos pensar que "lentos" modifica a la conjunción que le sigue, o sólo a "trenes". Por supuesto que una teoría adecuada descubriría la ambigüedad; una teoría de la verdad necesitaría en particular indicar cuándo una emisión de la oración podría ser verdadera bajo una interpretación y falsa bajo otra. Hasta aquí no hay dificultades para una teoría de la verdad. Pero Bar-Hillel hace la observación adicional de que el contexto de emisión podría resolver fácilmente la ambigüedad para todo hablante normal del castellano, y aun así la resolución podría depender de conocimientos generales en un modo que una teoría formal no podría (al menos prácticamente) capturar. Si reconocemos esto, como yo pienso que debemos hacer, aceptamos una limitación en lo que podríamos esperar de una teoría de la verdad. Dentro de esta limitación, todavía puede ser posible obtener una teoría que aprehenda un concepto de significado importante.

[6] Y. Bar-Hillel, *Languages and Information*, 182.

Hemos rozado ligeramente una pequeña parte de las consideraciones que han llevado a lingüistas y filósofos a dudar de la posibilidad de dar una teoría formal de la verdad para un lenguaje natural. He insinuado que el pesimismo es prematuro, particularmente en ausencia de una discusión de los criterios de adecuación. Por otra parte, sería tonto no reconocer una diferencia de intereses y métodos entre quienes estudian lenguajes inventados y quienes estudian lenguajes naturales.

Cuando lógicos y filósofos del lenguaje expresan sus reservas con respecto al tratamiento de los lenguajes naturales como sistemas formales, la razón podría ser su mayor interés en asuntos metateóricos como la consistencia, la completud, la decidibilidad. Estos estudios suponen un conocimiento exacto del lenguaje que se estudia, un tipo de precisión que sólo puede justificarse si los rasgos relevantes del lenguaje objeto se consideran fijados por una legislación. Esta actitud es sin duda inapropiada para el estudio empírico del lenguaje.

Sería engañoso, no obstante, concluir que hay dos tipos de lenguaje, el natural y el artificial. El contraste se delinea mejor en términos de los intereses rectores para ambos casos. Podemos buscar una descripción de la estructura de un lenguaje natural: la respuesta deberá ser una teoría empírica, abierta a la comprobación y sujeta a error, y condenada a ser hasta cierto punto incompleta y esquemática. O podemos buscar las propiedades formales de las estructuras que de esa manera abstraemos. Es la misma diferencia que hay entre la geometría aplicada y la pura.

He estado insistiendo en que no hay obstáculos definitivos en el camino hacia una teoría formal de la verdad para un lenguaje natural; todavía no está dicho por qué sería deseable. Las razones son de necesidad general y programática, pues lo que se impulsa no es una teoría particular sino un criterio de teorías. Se pretende que, de aceptarse este criterio, el estudio empírico del lenguaje ganará en claridad y significatividad. La cuestión de determinar si una teoría es correcta puede resolverse en una caracterización razonablemente nítida y comprobable; perseguimos teorías de un alto poder explicativo y predictivo, que empleen sofisticados recursos conceptuales que son ya bien conocidos. Entre los problemas que una teoría de la verdad satisfactoria resolvería, o ayudaría a resolver, figuran muchos que interesan tanto a lingüistas como a filósofos; así podemos anticipar el beneficio adicional de cierto grado de convergencia en los métodos e intereses de la filosofía y la lingüística. Trataré de iluminar brevemente estas observaciones.

Un requisito relativamente severo que se exige a una teoría de este tipo es que sea capaz de dar una caracterización recursiva de la oracionalidad. Esta parte de la teoría es comprobable al punto que tenemos, o podemos inventar, maneras confiables de determinar si una expresión es una oración. Imaginemos por el momento que podemos hacer esto lo suficientemente bien como para seguir adelante. Cuando definimos la oracionalidad estamos captando, a grandes rasgos, la idea de una expresión independientemente significativa. Pero la significatividad es sólo la sombra del significado; una teoría madura no sólo tendría que señalar

meramente las expresiones significativas, sino también dar sus significados. Esto lo reconocen hoy muchos lingüistas, pero la mayor parte de ellos admite no saber cómo satisfacer este requisito adicional de la teoría, o incluso no saber siquiera cómo formularlo.[7] Entonces me gustaría ahora decir algo más en favor de la opinión de que una teoría de la verdad debe "dar el significado" de las oraciones.

Una teoría de la verdad implica, para cada oración *s*, un enunciado de la forma "*s* es verdadero si y sólo si *p*" donde en el caso más simple "*p*" es reemplazado por *s*. Dado que las palabras "es verdadero si y sólo si" son invariantes, podemos interpretar, si queremos que significan "significa que". Así concebido, un ejemplo podría ser entonces: " 'Sócrates es sensato' significa que Sócrates es sensato".

Esta forma de destacar la relevancia de una teoría de la verdad en cuestiones de significado resulta esclarecedora, pero debemos tener en cuenta que puede fomentar ciertos errores. Uno de ellos es pensar que todo lo que podemos aprender acerca del significado de una oración particular a partir de una teoría de la verdad está contenido en el bicondicional que requiere la Convención T. Lo que pone de manifiesto aquello que podemos aprender es más bien la *prueba* de dicho bicondicional, pues la prueba debe demostrar, paso a paso, cómo el valor de la verdad de la oración depende de una estructura que se obtiene recursivamente. Una vez que tenemos una teoría, producir la prueba requerida es bastante fácil; el proceso podría mecanizarse.

Estudiar la estuctura de una oración a través de los ojos de una teoría de la verdad, es ver que está construida con un número finito de recursos, y que toda oración se satisface con cierto número finito de recursos; de esta manera, la estructura de la oración determina sus relaciones con otras oraciones. Y, por cierto, no sería posible dar las condiciones de verdad de todas las oraciones sin mostrar que algunas oraciones son consecuencias lógicas de otras; si consideramos que la estructura que revela una teoría de la verdad es gramática profunda, entonces la gramática y la lógica deben marchar codo con codo.

En un sentido, entonces, una teoría de la verdad describe la función que cada oración desempeña en el lenguaje, en tanto esa función dependa de que la oración sea una potencial portadora de verdad o de falsedad; y esa descripción se da en términos de estructura. Esta observación es, si duda, mucho menos clara que los hechos que la inspiran, pero al poner las cosas así me propongo justificar el argumento de que una teoría de la verdad muestra cómo "el significado de cada oración depende del significado de las palabras". O quizá sea suficiente decir que hemos dado un sentido a una idea sugestiva pero vaga; no hay razón para no dar la bienvenida a otras lecturas alternativas, si son igualmente claras. En todo caso, aceptar mi propuesta es dejar de lado el intento de hallar entidades que sirvan como significados de oraciones y palabras.

[7] Por ejemplo, véanse las observaciones de Chomsky sobre la semántica, "Topics in the Theory of Generative Grammar".

Una teoría de la verdad no las necesita; pero esto debe anotarse a su favor, al menos hasta tanto alguien proporcione una teoría del significado coherente y satisfactoria que emplee significados.

La Convención *T*, covenientemente modificada para hacerla aplicable a un lenguaje natural, provee un criterio de éxito para dar una descripción del significado. Pero, ¿cómo podemos probar esa descripción empíricamente? He aquí el segundo caso en el que podría engañarnos la aseveración de que los bicondicionales que requiere la Convención T podrían dar significados, pues esto sugiere erróneamente que la puesta a prueba de una teoría de la verdad exige una percepción directa de lo que significa cada oración. Pero de hecho, todo lo que hace falta es capacidad para reconocer si los bicondicionales que se requieren son verdaderos. Esto quiere decir que, en principio, probar la adecuación empírica de una teoría de la verdad no es más difícil que, para un hablante competente del castellano, decidir si oraciones como " 'La nieve es blanca' es verdadero si y sólo si la nieve es blanca" son verdaderas. Es así como la semántica o cuanto menos la teoría de la verdad, parece descansar sobre fundamentos empíricos tan firmes como los de la sintaxis. De hecho, en muchos casos podría resultar más fácil para un hablante decir cuáles son las condiciones de verdad de una oración que decir si la oración es gramática. Puede no resultar claro si "El niño parece dormir" es gramática; pero seguramente "El niño parece dormir" es verdadera si y sólo si el niño parece dormir.

He estado imaginando la situación en la cual el metalenguaje contiene al lenguaje objeto, donde podemos aspirar a que un hablante nativo reaccione a los bicondicionales que le son familiares y que conectan una oración y su descripción. Estaremos ante un caso más extremo si queremos someter a prueba una teoría formulada en nuestro propio lenguaje acerca del lenguaje de un hablante extranjero. Aquí puede ponerse nuevamente a prueba una teoría de la verdad, si bien no tan fácil o directamente como antes. El proceso deberá tener un parecido con el que describiera Quine en el capítulo 2 de *Word and Object*. Descubriremos condiciones bajo las cuales el hablante extranjero concordará o disentirá con una variedad de sus oraciones. Consideraremos que las condiciones relevantes son las condiciones de verdad de sus oraciones. Tendremos que suponer que para los casos simples u obvios la mayoría de sus asensos corresponden a oraciones verdaderas y la mayoría de sus disensos corresponden a oraciones falsas: una suposición inevitable dado que cualquier alternativa es ininteligible. Pero Quine está en los cierto, creo yo, cuando sostiene que luego de que toda la evidencia haya sido registrada permanecerá un grado importante de indeterminación; cierto número de teorías de la verdad significativamente diferentes entre sí responderá igualmente bien a la evidencia.[8]

[8] Para mayor discusión de los principios que guían la interpretación véanse los Ensayos 2, 9 a 11, 13 y 14. Hay algunas diferencias entre el método de traducción radical de Quine y el método que sugiero; véase el Ensayo 16.

Dar una consideración sistemática de la verdad que resulte esencial para la semántica empírica es, en cierta forma, una simple cuestión de enunciar viejas metas de una manera más nítida. Pero la línea divisoria entre clarificación e innovación en ciencia sigue siendo borrosa, y es probable que el cambio traiga consigo un trastocamiento de prioridades en la investigación lingüística. Se desdibujaría la importancia de algunos problemas que han dominado en los recientes trabajos de semántica, como el intento de dar "el significado" de las oraciones y de explicar la sinonimia, la analiticidad y la ambigüedad. Para el primero de ello, la teoría de la verdad provee una especie de sustituto; el segundo y el tercero se tornan apéndices innecesarios; el cuarto reaparece bajo una forma especial. Pero también surgirían problemas de fondo, como dificultades para la referencia, para dar una semántica satisfactoria para oraciones modales, oraciones acerca de actitudes proposicionales, términos masa, modificación adverbial, adjetivos atributivos, imperativos e interrogativos; y así sucesivamente, en una larga lista que, en su mayor parte, a los filósofos les es familiar.[9]

Hay una duda abierta en torno del grado de realineamiento que podemos esperar en lingüística. Esto depende en gran medida de hasta qué punto pueda identificarse la estructura que una teoría revela con la estructura profunda que persiguen los gramáticos transformacionales. En un aspecto, la estructura lógica (que es como podemos llamar a la estructura revelada por una teoría de la verdad) y la estructura profunda podrían ser una misma, pues de ambas se pretende que sean el fundamento de la semántica. La estructura profunda debe servir también, sin embargo, como base para las transformaciones que producen estructuras de superficie, y no sabemos si la estructura lógica puede cumplir esta tarea, o cumplirla bien.[10]

Por último, algunos lingüistas pretenden que la estructura profunda refleja la "gramática internalizada" de los hablantes del lenguaje. Chomsky en particular ha sostenido que la superioridad de las gramáticas transformacionales sobre otras que podrían ser igualmente buenas para explicar la totalidad de las oraciones gramáticas reside en el hecho de que las gramáticas transformacionales pueden hacerse "corresponder a la intuición lingüística del hablante nativo".[11] El problema es hallar un test relativamente claro para determinar cuándo una teoría corresponde a la intuición lingüística de un hablante. Quisiera sugerir que si consideramos que la estructura profunda tiene forma lógica podremos tener una especie de aproximación empírica a esta idea. Veamos un pasaje de Chomsky sobre el tema.[12]

[9] Ha habido muchos cambios desde que esto se escribió: los lingüistas han reconocido estos problemas y muchos otros, y han acicateado a los filósofos para que éstos aprecien problemas que nunca habían notado. (Nota agregada en 1982.)

[10] Para el desarrollo de este tema, véase G. Harman, "Logical Form".

[11] N. Chomsky, *Aspects of the Theory of Syntax*, 24.

[12] *Ibídem*, 22.

Chomsky dice que si bien las dos oraciones siguientes tienen la misma estructura de superficie, difieren en su estructura profunda:

(1) *I Persuaded John to leave*. [Persuadí a Juan de que se fuera.]

(2) I *expected John to leave*. [Esperé que Juan se fuera.]

La demostración descansa principalmente sobre la observación de que cuando transformamos a la voz pasiva una oración anidada en una oración parecida a (2), el resultado es "cognitivamente sinónimo" de la forma activa; pero una transformación similar no redunda en un resultado sinónimo para la análoga de (1). La observación es sin duda correcta, pero ¿cómo hace para demostrar que (1) y (2) tienen estructuras profundas radicalmente diferentes? A lo sumo la evidencia sugiere que una teoría que asigna diferentes estructuras a (1) y a (2) puede ser más simple que una que no lo hace. Pero no queda claro cómo se han explotado nuestras intuiciones lingüísticas para notar aquí una diferencia.

Pero desde luego que Chomsky tiene razón; hay un contraste entre (1) y (2), y éste emerge con toda su fuerza en el momento en que comenzamos a pensar en la construcción de una teoría de la verdad. Tanto es así que basta con preguntarse por el rol semántico de la palabra "John" en ambas oraciones. En (1), "John" puede reemplazarse por cualquier término correferente sin alterar el valor de verdad de (1); esto no ocurre con (2). La contribución de la palabra "John" a las condiciones de verdad de (1) debe ser entonces totalmente diferente de su contribución a las condiciones de verdad de (2). Esta forma de expresar que hay una diferencia en la estructura semántica de (1) y de (2) no requiere apelar al "conocimiento tácito" de la gramática por parte del hablante ni a la "competencia intrínseca del hablante nativo idealizado". Descansa en el conocimiento explícito que todo hablante del inglés tiene de la forma en que (1) y (2) pueden variar su verdad en caso de sustituciones de la palabra "John".

Pero estas últimas observacioes no comienzan aún a hacer justicia al método de verdad. Ellas manifiestan que al tener presentes los requisitos de una teoría de la verdad podemos corporizar nuestra percepción de una diferencia de estructura entre (1) y (2). Hasta el momento, sin embargo, estamos apelando a una evidencia de tipo muy parecido a la que utiliza Chomsky: principalmente cuestiones de pérdida o preservación del valor de verdad en casos de transformaciones. Estas consideraciones, sin duda, continuarán guiando las tareas constructivas y analíticas de los lingüistas tal como lo han hecho largamente con los filósofos. La belleza de una teoría del tipo que hemos estado discutiendo está en que estos indicios de estructura, por útiles o esenciales que puedan ser para el descubrimiento de una teoría adecuada, no requieren desempeñar ninguna función directa en la puesta a prueba del producto final.

5

En defensa de la Convención T

Supongamos que alguien dice: "Hay un millón de estrellas esta noche", y otro le contesta: "Eso es verdad", entonces nada más manifiesto que lo que el primero dijo es verdadero si y sólo si lo que dijo el otro es verdadero. Este familiar efecto se debe a la reciprocidad de dos recursos, uno consistente en una forma de referir expresiones (tarea que aquí cumple el demostrativo "eso"), el otro el concepto de verdad. El primer recurso nos lleva del discurso del mundo al discurso del lenguaje; el segundo nos trae de vuelta otra vez.

Hemos aprendido a representar estos hechos con oraciones de la forma "La oración 'Hay un millón de estrellas esta noche' es verdadera si y sólo si hay un millón de estrellas esta noche". Dado que las oraciones T (como podemos llamarlas) son tan obviamente verdaderas, algunos filósofos han pensado que el concepto de verdad, al menos en su aplicación a oraciones, es trivial. Pero esto indudablemente es un error, tal como Ramsey lo señaló quizá por primera vez, pues las oraciones T proveen una alternativa para hablar de la verdad sólo para algunos contextos. Las oraciones T no nos serán útiles si buscamos equivalentes de "Toda oración que decía Aristóteles era falsa" o de "Lo que dijiste el martes pasado era verdad".[1]

En consecuencia, las oraciones T no enseñan a vivir sin predicado de verdad; pero tomados en conjunto, sí nos muestran cómo sería tener uno. Pues, dado que hay una oración T correspondiente a cada oración del lenguaje del cual se cuestiona la verdad, la totalidad de las oraciones T fija exactamente la extensión, entre las oraciones, de todo predicado que juega el papel de las palabras "es verdadero". De esto resulta claro que si bien las oraciones T no definen la verdad, pueden usarse para definir la predicación de verdad: todo predicado es un predicado de verdad que hace verdaderas a todas las oraciones T.

Las oraciones T iluminan la verdad, sea cual fuere el grado en que lo hagan, por medio del agotamiento de los casos, pero dado que para todo lenguaje interesante el número de casos es infinito, resulta instructivo averiguar cuáles son las propiedades que debe tener la verdad para poder actuar como las oraciones T dicen que lo hace. Y hacer claras y

[1] F. P. Ramsey, "Facts and Propositions", 143.

explícitas las propiedades de una manera instructiva, digamos, finita, equivale a proveer una teoría de la verdad. Podemos pensar que una teoría de la verdad para un lenguaje L es simplemente una oración T que contiene un predicado *t* tal que T tiene por consecuencias lógicas a todas las oraciones de la forma "*s* es verdadero si y sólo si *p*" donde "*s*" es reemplazada por una descripción canónica de una oración de L, "*p*" es reemplazada por esa oración (o su traducción), y "es verdadero" es reemplazado, en caso necesario, por *t*.

Desde luego, esto es esencialmente la Convención T de Tarski.[2] Sólo difiere en un aspecto significativo: no requiere que las condiciones del predicado de verdad basten para una definición explícita. (Volveré a este punto.) A través de los escritos de Tarski sobre la verdad, la Convención T proporciona la transición de la intuición informal en torno al concepto de verdad a una manifestación definida del problema. Encontramos por primera vez a la Convención T en "The Concept of Truth in Formalized Languages" (publicado en polaco en 1933, en alemán en 1936, y en inglés en 1956). En un artículo posterior destinado principalmente a los filósofos, Tarski apela a la oraciones T para relacionar temas filosóficos clásicos con su obra sobre la semántica.[3] Retoma más tarde esta idea en otro popular trabajo.[4]

Creo que, a pesar de toda su familiaridad, los filósofos han pasado por alto un rasgo principal de la Convención T. Es el siguiente: la Convención T y las oraciones T proveen el único contacto entre las verdades intuitivamente obvias acerca de la verdad y la semántica formal. Sin la Convención T, no tendríamos razones para creer que la verdad es lo que Tarski nos ha enseñado cómo caracterizar.

El peso que Tarski atribuye a la Convención T es interesante también en otro sentido, pues genera una ligera tensión terminológica. Según Tarski:

> La *semántica* es una disciplina que, en términos sencillos, se ocupa de ciertas relaciones entre las expresiones de un lenguaje y los objetos... "referidos" por dichas expresiones.[5]

Da como ejemplos de conceptos semánticos la *designación* (denotación), que relaciona a los términos singulares con lo que ellos denotan; la *satisfacción*, que se aplica entre una oración abierta y la entidad o entidades para las que es verdadera; y la *definición*, que relaciona una ecuación y un número que ella determina exclusivamente. Continúa Tarski:

> Mientras las palabras "*designa*", "*satisface*" y "*define*" expresan relaciones..., la palabra "*verdadero*" es de una naturaleza lógica diferente: ex-

[2] A. Tarski, "The Concept of Truth in Formalized Languages", 187, 188.
[3] A. Tarski, "The Semantic Conception of Truth".
[4] A. Tarski, "Truth and Proof".
[5] A. Tarski, "The Semantic Conception of Truth", 345.

presa una propiedad (o denota una clase) de ciertas expresiones, cuales son las oraciones.[6]

Pasa a señalar que "la forma más simple y natural" de definir la verdad es por medio de la satisfacción, y esto explica que este autor considere la verdad como un concepto semántico. Pero no podemos dejar de notar que, a través de la propia explicación de Tarski, la verdad no resulta ser un concepto semántico. Una mirada a la gramática lógica de las oraciones T muestra que es esencial que el predicado de verdad no exprese una relación: si lo hiciera, no podría producirse en el lado derecho del bicondicional de la oración T la fundamental desaparición de todos los conceptos semánticos, y en general de todo lo que no sea la oración misma cuyas condiciones de verdad aquél establece (o una traducción de dicha oración).[7]

Mucho se avanzaría en la justificación de la terminología de Tarski si pudiera demostrarse que toda teoría que satisface el criterio de la Convención T tiene recursos adecuados para caracterizar la relación de satisfacción. Indudablemente hay lenguajes para los cuales esto no vale, pero sería interesante que pudiera hacerlo para lenguajes lo suficientemente ricos. Trabajos recientes sugieren una respuesta positiva.[8] La sensación de que la verdad pertenece a la clase de los otros conceptos semánticos también se ve reforzada por el hecho de que en cada caso hay un paradigma que tanto usa como menciona a la misma expresión: "Platón" designa a Platón; y satisface a "x es blanco" si y sólo si y es blanco; "la nieve es blanca" es verdadero si y sólo si la nieve es blanca; la ecuación "$2x = 1$" define al número y si y sólo si $2y = 1$, etcétera. La existencia de tal paradigma, sin embargo, no puede tomarse como un signo de lo semántico, pues en un modelo no hay un paradigma análogo apropiado para conceptos semánticos relativizados como la verdad.

Mi propósito presente es defender, en base a su interés e importancia filosóficos, a los conceptos que satisfacen el criterio de la Convención T (y para los cuales hay, entonces, un paradigma familiar). Tarski apeló a la Convención T, entre otras razones, porque esperaba persuadir a los filósofos de que el trabajo formal en semántica era apropiado para sus fines. Resulta escandaloso que esta causa todavía necesite defensa en ciertos círculos; pero no es esto lo que me preocupa en este trabajo. Me dirijo ahora a quienes aprecian por entero la relevancia de la semántica formal en los temas filosóficos, pero no ven diferencias significativas en-

[6] *Ibídem*, 345.

[7] Esta afirmación se modifica más adelante.

[8] Véase J. Wallace, "On the Frame of Reference", y "Convention T and Substitutional Quantification"; también L. Tharp, "Truth, Quantification, and Abstract Objects". Si bien Kripke critica algunos de los argumentos de Wallace y Tharp, coincide con la conclusión de que en un lenguaje con poder expresivo normal, al menos algunos de los cuantificadores deben ser provistos con el tipo de semántica para la cual está concebida la satisfacción de Tarski. Véase S. Kripke, "Is There a Problem About Substitutional Quantification?" (Nota agregada en 1982.)

tre teorías que concuerdan con la Convención T, y teorías que no lo hacen. (Incluso existe el peligro de que los ignorantes y los expertos reúnan sus fuerzas; los primeros, al escuchar rumores de mundos posibles, líneas hereditarias de transmundo, contrapartes, y otros, están en situación de pensar que *ahora* la semántica está yendo hacia alguna parte: fuera de este mundo, en todo caso.)

Antes de exponer las razones por las que pienso que la Convención T aprehende un concepto que vale la pena estudiar, quiero registrar algunos renunciamientos.

Primero, perseguir una teoría acorde con la Convención T no es, en sí mismo al menos, estatuir una lógica o una semántica del Modelo T. La Convención T, en la forma esquemática en que la he dado, no hace mención de extensionalidad, funcionalidad veritativa, o lógica de primer orden. Nos invita a utilizar cuanto recurso podamos idear en forma apropiada para salvar la brecha entre oración mencionada y oración usada. Toda restricción de ontología, ideología o poder inferencial sólo encuentra eco favorable, desde el presente punto de vista, si resulta de adoptar la Convención T como piedra de toque. Quiero defender a la Convención como criterio de teorías, y no a teorías particulares que hayan demostrado satisfacer a la Convención en casos particulares, o a los recursos a los cuales ellas hayan podido limitarse.

La Convención T define una meta irrelevante para buena parte del trabajo contemporáneo en semántica. Muchas teorías que caracterizan o definen un concepto de verdad relativizado (verdad en un modelo, verdad en una interpretación, valuación, o mundo posible) se dirigen desde un principio en una dirección distinta a la que propone la Convención T. Dado que sustituyen el predicado de verdad monádico de las oraciones T por un concepto relacional, tales teorías no pueden transitar hasta el final el último paso de la recursión respecto de la verdad, o satisfacción, que es esencial para la propiedad de relevamiento de citas de las oraciones T. Según una tradición que inició el propio Tarski, se llama teoría general de la verdad a una teoría relativizada de la cual la teoría absoluta (que satisface la Convención T) es un caso especial. Tarski dice que "el concepto de *oración correcta o verdadera en un dominio individual* [es] un concepto de carácter relativo [que] desempeña un papel mucho más importante que el concepto absoluto de verdad e incluye a éste como un caso especial".[9] Desde luego, queda perfectamente claro el sentido en que esto es correcto. Por otra parte, es importante recordar que las oraciones T no se dan como teoremas de una teoría relativizada de la verdad, y por ello una teoría de ese tipo no necesariamente reviste el mismo interés filosófico que una teoría que satisface la Convención T.

Espero quede claro que con estas observaciones no pretendo sugerir que las teorías que no producen oraciones T como teoremas tengan algo de malo, defectuoso o engañoso. Por el contrario, es obvio que esas teorías iluminan importantes conceptos (tales como completud y conse-

[9] A. Tarski, "The Concept of Truth in Formalized Languages", 199.

cuencia lógica); realmente no necesito extenderme sobre este tema en el contexto presente. Mi tesis solamente trata de que hay importantes diferencias entre teorías de la verdad relativa y de la verdad absoluta, y de que las diferencias hacen que las teorías de los dos tipos sean apropiadas como respuestas a cuestiones diferentes. En particular, los filósofos del lenguaje tienen razones para interesarse más por teorías que satisfacen la Convención T.

Por lo general, las teorías de la verdad que se basan en la interpretación sustitucional de la cuantificación no producen las oraciones T que requiere la Convención T. Hay excepciones en el caso de lenguajes objeto cuyas oraciones atómicas verdaderas pueden proporcionarse efectivamente, pero las excepciones no pueden incluir lenguajes de un poder expresivo similar en interés al de los lenguajes naturales. A diferencia de las teorías de la verdad relativa, las teorías sustitucionales carecen de una virtud evidente que compense su falta de satisfacción de la Convención T.[10]

Por último, mi tesis no dice que *todo* lo que debamos exigir a una teoría semántica sea que cumpla las normas de la Convención T. Me atrevo a sugerir que las teorías que satisfacen la Convención T son capaces de explicar más de lo que a menudo se piensa. Pero puede haber casos de elección entre ese tipo de teorías en base a criterios futuros; y naturalmente hay muchas cosas que queremos saber que quedan fuera de su campo de acción.

Veamos ahora un catálogo de algunos de los rasgos de las teorías que concuerdan con la Convención T y que deberían inclinar a los filósofos del lenguaje hacia su uso.

El mérito principal de la Convención T es sustituir un problema importante pero sombrío por una tarea cuyo objetivo es claro. Luego de la sustitución uno aprecia mejor lo que se buscaba en primer término y se adentra en la etiología de la confusión. La pregunta original no resulta confusa, sino solamente vaga. Esta es: ¿qué significa para una oración (o emisión o enunciado) ser verdadera? La confusión nos amenaza cuando reformulamos esta pregunta como ¿qué hace verdadera a una oración? El problema real aparece cuando a su vez entendemos que esto sugiere que la verdad debe explicarse en términos de una relación entre una oración interpretada como un todo y alguna entidad, quizás un hecho, o un estado de cosas. La Convención T expresa la forma de hacer la pregunta original sin dar lugar a estas formulaciones subsiguientes. La forma de las oraciones T ya insinúa que una teoría puede caracterizar la propiedad de verdad sin tener que hallar entidades a las cuales correspondan diferencialmente oraciones que tienen esa propiedad. (No quiero decir que una teoría absoluta de la verdad no sea en cierto sentido una

[10] Si bien la conclusión general de estas observaciones puede ser acertada, éstas son poco cuidadosas. En particular, pasé por alto la consideración de lenguaje con cuantificadores tanto sustitucionales como ónticos. (Véase la nota 8 más atrás). (Nota agregada en 1982.)

"teoría de correspondencia" de la verdad. Pero las entidades que ella invoca son oraciones, algo muy diferente de hechos o estados de cosas.)[11]

Una teoría que satisface la Convención T tiene, entonces, la virtud de ser una respuesta a una buena pregunta. El enunciado de la pregunta es tan persuasivo, en relación al concepto intuitivo de verdad, como encontramos que lo es la verdad de las oraciones T. Tanto el enunciado de la pregunta, como el carácter de las respuestas que admite, revelan que no se usan nociones semánticas que no estén enteramente caracterizadas en sus aspectos relevantes, sintácticamente por la Convención T y materialmente por las teorías que ésta acepta.

Una teoría recursiva de la verdad absoluta del tipo exigido por la Convención T también provee una respuesta, a primera vista *per accidens*, a un problema muy distinto. Este problema puede expresarse como el de mostrar o explicar de qué manera el significado de una oración depende del significado de sus partes. Una teoría de la verdad absoluta da una respuesta en el sentido siguiente. Dado que hay una infinidad de oraciones T a tenerse en cuenta, la teoría debe operar seleccionando un número finito de expresiones veritativas relevantes y un número finito de construcciones que afectan a la verdad a partir de las cuales se componen todas las oraciones. A continuación la teoría proporciona abiertamente las propiedades semánticas de algunas de las expresiones básicas y explica de qué modo las construcciones afectan las propiedades semánticas de las expresiones sobre las cuales operan.

En el párrafo previo, la noción de significado a que se hace referencia en la consigna "El significado de la oración depende del significado de sus partes" no consiste, desde luego, en la noción que opone significado a referencia, ni en una noción que suponga que los significados son entidades. Esa consigna refleja una importante verdad, una verdad a la cual, a mi entender, una teoría de la verdad *confiere* contenido claro. El hecho de que lo haga sin introducir significados a modo de entidades es una de sus cualidades más gratificantes.

Las teorías de la verdad absoluta proveen necesariamente un análisis de la estructura relevante para la verdad y la inferencia. Estas teorías ofrecen así una respuesta no trivial a la cuestión de qué implica ser la forma lógica de una oración. Una teoría de la verdad no ofrece una definición de consecuencia lógica o de verdad lógica, pero proveerá la evidencia de que ciertas oraciones son verdaderas únicamente sobre la base de las propiedades asignadas a las constantes lógicas. Las constantes lógicas pueden identificarse como aquellos rasgos iterativos del lenguaje que requieren una cláusula recursiva en la caracterización de la verdad o de la satisfacción. En esta consideración la forma lógica será, desde luego, relativa a la elección de un metalenguaje (con su lógica) y de una teoría de la verdad.

Todo lo que hasta ahora he dicho en defensa de la Convención T es que se la puede entender como una formulación clara de varios asuntos

[11] Véase el Ensayo 3.

problemáticos de la filosofía del lenguaje, y que así entendida invita a solucionarse que gozan de méritos conspicuos. Pero, ¿por qué tendríamos que verla así? ¿No puede haber una convención rival que formule las cosas mejor? Nos ocupamos ahora de estas cuestiones.

El principal, si no el único interés último de la filosofía del lenguaje es la comprensión de los lenguajes naturales. Hay muchos argumentos para restringir la palabra "lenguaje" a los sistemas de signos que son o han sido de uso real: los sistemas formales no interpretados no son lenguajes a raíz de su falta de significado, mientras que los sistemas formales interpretados se consideran más bien extensiones o fragmentos de los lenguajes naturales que les prestan vida. La meta inevitable de la teoría semántica es una teoría de un lenguaje natural que descanse en un lenguaje natural (el mismo u otro). Pero como lo señalara Tarski, podemos suponer casi hasta la obviedad que el empeño en esta meta conduce a una paradoja. Este autor escribió:

> Un rasgo característico del lenguaje coloquial... es su universalidad. No armonizaría con el espíritu de este lenguaje la existencia de una palabra en algún otro lenguaje que no pudiera ser traducida a él; podría sostenerse que "si podemos hablar acerca de cualquier cosa, también podemos hablar de ello en lenguaje coloquial". Si es que vamos a seguir sosteniendo esta universalidad del lenguaje cotidiano en relación con las investigaciones semánticas, para mantener una coherencia debemos admitir también dentro del lenguaje, además de sus oraciones y otras expresiones, los nombres de estas oraciones y expresiones, y las oraciones que contienen estos nombres, al igual que expresiones semánticas tales como "oración verdadera"...[12]

Una vez que se ha admitido todo esto dentro del lenguaje, surgen las antinomias semánticas. El ideal de una teoría de la verdad para un lenguaje natural en un lenguaje natural es entonces inalcanzable si nos limitamos a los métodos de Tarski. Luego surge la cuestión de cómo claudicar lo menos posible, y es aquí donde las teorías permitidas por la Convención T parecen óptimas desde muchos aspectos. Estas teorías pueden conferir a un predicado de verdad las propiedades requeridas sin explotar recursos conceptuales que no formen parte del lenguaje para el cual el predicado es aplicable. Lo único que puede no estar en el lenguaje objeto es el propio predicado de verdad (y el predicado de satisfacción). Aquí resulta esencial requerir nada más que una *teoría* de la verdad; ir más allá de eso hacia una definición explícita ensancharía la brecha entre los recursos del lenguaje objeto y los del metalenguaje.[13] Pero si buscamos nada más que una teoría del tipo que hemos estado discutiendo, entonces la ontología del metalenguaje puede ser la misma

[12] A. Tarski, "The Concept of Truth in Formalized Languages", 164.

[13] Para un ejemplo instructivo, véase W. V. Quine, "On an Application of Tarski's Theory of Truth".

que la del lenguaje objeto, y el incremento de ideología puede verse limitado a los conceptos semánticos.

Tarski está en lo cierto, creo yo, cuando propone que entendamos a los lenguajes naturales como esencialmente intertraducibles (aunque no veo por qué esto deba requerir una traducción palabra por palabra). La propuesta idealiza la flexibilidad y expandibilidad de los lenguajes naturales, pero puede justificarse mediante un argumento trascendental (que no explicaré aquí).[14] La Convención T exige que cada oración del lenguaje objeto tenga una traducción en el metalenguaje. La presión en la dirección opuesta deriva del deseo de dar cuenta, en lo posible, de los recursos lingüísticos que dominamos.

Existen teorías expresadas en un metalenguaje enteramente extensional que intentan iluminar los rasgos semánticos de un lenguaje objeto intensional. El interés de tales teorías depende de la presunción de que el lenguaje objeto refleja rasgos importantes de un lenguaje natural. Pero entonces aparece la pregunta de cómo entendemos al metalenguaje. Pues en un sentido claro el metalenguaje excede al poder expresivo del lenguaje objeto. Para cada oración "intensional" del lenguaje objeto hay una oración extensional correspondiente (acerca de mundos posibles, contrapartes, etcétera) con las mismas condiciones de verdad en el metalenguaje; pero debe haber oraciones metalingüísticas que intuitivamente tienen el mismo tema sujeto y que no tienen oraciones correspondientes en el lenguaje objeto. Esas teorías tratan a la intensionalidad como carencia de poder expresivo; una carencia que la misma teoría nos enseña a compensar.[15] Pero si comprendemos nuestro metalenguaje, estamos usando un sistema de conceptos y un lenguaje que es aquél para el cual *realmente* queremos una teoría, pues este sistema, más rico, es nuestro sistema natural. Y afortunadamente este sistema más rico no genera dificultades para una teoría de la verdad que satisface la Convención T, pues es extensional. (No considero aquí si realmente entendemos esos metalenguajes.)

Si una teoría semántica asegura ser aplicable, aunque sea esquemáticamente, a un lenguaje natural, entonces debe ser de carácter empírico y sujeta a comprobación. Quisiera esbozar en estas páginas finales mis razones para pensar que una teoría que satisface la Convención T es verificable en una forma muy interesante.

Por supuesto, no se trata una teoría de la verdad como si fuera empírica si su adecuación se juzga solamente en términos de las oraciones T que implica, y las oraciones T se verifican sólo por su forma; esto ocurre si *suponemos* que el lenguaje objeto está contenido en el metalenguaje. Cuando esta suposición se mitiga, la teoría puede tornarse empírica. Y esto se cumple en el momento en que se anuncia que la teoría es aplicable al habla de una persona o grupo particular. El hecho de que el lenguaje objeto esté contenido en el metalenguaje no impide a la teoría

[14] Pero véanse los Ensayos 13 y 14.
[15] Debo esta explicación a John Wallace.

tener contenido empírico: más bien este hecho puede calificarse como el hecho a ser verificado.

En los casos que cuentan, entonces, no podemos suponer que los lenguajes descritos y descritores coinciden. Por cierto, tampoco podemos usar un criterio formal de traducción sin suponer establecida la cuestión de la aplicación empírica. Pero entonces, ¿qué pasa con la Convención T? ¿Cómo se reconocerá a una oración T, por no hablar de reconocerla como verdadera?

Sugiero que puede bastar con requerir que las oraciones T sean verdaderas. Está claro que con esto sólo alcanza para determinar correctamente la extensión del predicado de verdad. Si consideramos una oración T cualquiera, esta proposición solamente requiere que si una oración verdadera es descrita como verdadera, luego sus condiciones de verdad están dadas por alguna oración verdadera. Pero cuando consideramos la necesidad compulsiva de comparar verdad con verdad a través de todo el lenguaje, nos damos cuenta de que toda teoría aceptable según esta norma debe ofrecer efectivamente un manual de traducción utilizable del lenguaje objeto al metalenguaje. El efecto deseado es normal en la construcción de teorías: extraer un concepto rico (aquí algo razonablemente próximo a la traducción) a partir de pequeñas partículas de evidencia (aquí los valores de verdad de las oraciones) mediante la imposición de una estructura formal sobre una cantidad suficiente de partículas. Si caracterizamos a las oraciones T solamente por su forma, como lo hizo Tarski, es posible, usando los métodos de Tarski, definir la verdad sin usar conceptos semánticos. Si consideramos a las oraciones T verificables, luego una teoría de la verdad expresa de qué modos podemos ir de la verdad hacia algo parecido al significado; lo suficientemente parecido al significado de manera que si alguien pudiera verificar una teoría para un lenguaje en la forma que yo propongo, sería capaz de utilizar ese lenguaje para la comunicación.[16]

¿Qué hace plausible que el rango de teorías aceptables según esta norma excluya intuitivamente a teorías sin valor? Uno de los méritos principales de esta aproximación, cual es que toca lo observable sólo cuando alcanza a las oraciones, parece dejar el análisis de la estructura interna simplemente fuera del alcance. La articulación de oraciones en términos singulares, cuantificadores, predicados, conectivos, así como la vinculación de expresiones con entidades en la caracterización de la satisfacción, debe ser considerada como un exceso de construcción teórica, a comprobarse sólo a través de su éxito en la predicción de las condiciones de verdad de las oraciones.[17]

Un factor importante y ciertamente esencial para convertir a una teoría de la verdad en una teoría creíble de la interpretación es la relativización al hablante y al tiempo. Cuando están presentes elementos deícticos o demostrativos, no se puede hablar de oraciones que son verda-

[16] Para detalles y limitaciones véanse los Ensayos 9-12.
[17] Este punto está más desarrollado en los Ensayos 15 y 16.

deras o falsas, sino sólo de oraciones relativas a un hablante y a un tiempo. Podemos tomar alternativamente a la verdad como una propiedad, no de las oraciones, sino de las emisiones o de los actos de habla. Estos a su vez pueden identificarse con ciertos triplos ordenados de oraciones, tiempos y hablantes. La elección entre estas dos alternativas involucra algunos problemas sutiles, pero por el momento tengo que dejarlos de lado. El punto ahora es que una teoría que determine como verdaderas a las oraciones correctas en los tiempos correctos y para los hablantes correctos estará mucho más próxima a una teoría que interpreta correctamente las oraciones que una que puede ignorar los parámetros adicionales. En efecto, podremos saber algo directamente relacionado con la extensión de una oración abierta "F*x*" cuando conozcamos las condiciones de verdad de "Eso es una *x* tal que F*x*".

La oración metalingüística que da las condiciones de verdad (relativizadas) de una oración con deícticos no puede tener la forma de una oración T, aunque abandonemos el requerimiento de un criterio puramente sintáctico para la relación entre oración descrita y oración usada. El problema es que las variables que tratan de personas y tiempos, las variables introducidas para dar lugar a la relativización, deben aparecer en el enunciado de las condiciones de verdad. La verdad relativa a un tiempo y a una persona parecería encontrarse en las mismas condiciones que la verdad en un modelo.

De ello se desprende que la interesante diferencia entre una teoría de la verdad en un modelo y una teoría de la verdad "absoluta" no puede describirse, como lo he estado intentando, diciendo que en la segunda, pero no en la primera, la verdad está caracterizada para cada oración *s* sin apelar a recursos que no estén en *s* misma. La razón, como acabamos de verlo, es que la verdad "absoluta" se torna relativa cuando se la aplica a un lenguaje natural. Sin embargo permanece una diferencia, a la que yo adjudicaría mucha importancia. La verificación de las instancias de las oraciones T, o más bien de sus sustitutas en una teoría relativizada a hablantes y a tiempos, continúa siendo respetablemente empírica. Sin duda entrará en juego algún concepto pragmático de *demostración* entre hablantes, tiempos y objetos. Pero podemos esperar explicar tal concepto sin recurrir a nociones como verdad, significado, sinonimia o traducción. No puede decirse lo mismo de la verdad en un modelo. La Convención T, aun cuando deba torcerse para adecuarse a las torpes forma del lenguaje natural, señala el camino hacia una teoría radical de la interpretación.

Aplicaciones

6

La cita

La cita es un recurso usado para referir formas tipográficas o fonéticas mediante la exhibición de muestras, esto es, inscripciones o emisiones que tienen esas formas. Esta caracterización es amplia y vaga: lo suficientemente amplia como para incluir no sólo las comillas escritas y las frases dichas como "y yo cito", sino también las citas a dedo blandiente que a menudo usan los filósofos condenados a leer en voz alta lo que han escrito; y lo suficientemente vaga como para dejar abierta la cuestión de si las palabras que dieron comienzo a esta oración ("Esta caracterización") muestran una forma de cita.

En la cita, el lenguaje no sólo se vuelve hacia sí mismo sino que lo hace palabra por palabra y expresión por expresión, y este giro reflexivo es inseparable de la conveniencia y aplicabilidad universal del recurso. Con esto ya tenemos lo suficiente como para llamar la atención del filósofo del lenguaje; pero discernimos también conexiones con otras áreas de interés como oraciones acerca de actitudes proposicionales, ejecutores explícitos, y teorías de ilustración de la referencia. Si los problemas que genera la cita parecen triviales por comparación, podemos saludar el hallazgo de una entrada franca a este laberinto.

Cuando fui iniciado en los misterios de la lógica y la semántica, se presentaba usualmente la cita como un recurso algo sombrío, y la presentación se acompañaba de un severo sermón sobre el pecado de confundir el uso y la mención de expresiones. La conexión entre la cita por una parte y la distinción uso-mención por la otra es obvia, pues una expresión que sería usada si una de sus instancias particulares apareciera en un contexto normal es mencionada si una de sus instancias particulares aparece entre comillas (o alguna indicación similar de la cita). La invitación al pecado se observa tal vez en la facilidad con que las comillas pueden pasar desapercibidas u omitirse. Pero otras críticas más profundas de la cita pintan un cuadro más grave. Así Tarski, en "The Concept of Truth in Formalized Languages", examina las posibilidades de una teoría articulada de las comillas, y decide que uno se ve rápidamente desviado hacia absurdos, ambigüedades y contradicciones.[1] Quine escribe en *Mathematical Logic*: "Un uso escrupuloso de las

[1] A. Tarski, "The Concept of Truth in Formalized Languages", 159-62.

comillas es la principal medida práctica para evitar confundir los objetos con sus nombres...", pero luego agrega que la cita:

> ...tiene cierta característica anómala que requiere especial cautela: desde el punto de vista del análisis lógico, cada cita completa debe entenderse como una única palabra o signo, cuyas partes no cuentan más que los *serifs* o las sílabas. Una cita no es una *descripción* sino un *jeroglífico*; designa a su objeto no sólo describiéndolo en términos de otros objetos, sino también ilustrándolo. El significado del todo no depende de los significados de las palabras constituyentes.[2]

Y Church, mientras elogia a Frege por su cuidadoso empleo de la cita para no caer en equivocaciones, declara evitar la cita por ser "confusa", "torpe en la práctica... y expuesta a desafortunados abusos y malinterpretaciones".[3] Hay más de un atisbo, entonces, acerca de algo oscuro o confuso en torno de la cita. Pero esto no puede ser cierto. No hay nada de malo en el recurso mismo. Inadecuadas y confusas son nuestras teorías acerca de su funcionamiento.

Se dice a menudo que en la cita, las expresiones citadas son mencionadas y no usadas. La primera parte de esta afirmación es relativamente clara. La segunda parte, la que dice que las expresiones citadas no son usadas, parece sospechosa. ¿Por qué la incorporación a una cita no es un uso de una expresión? Una respuesta admisible sería que, por supuesto, en *algún* sentido el material citado se usa pero su uso en la cita no se relaciona con su *significado* en el lenguaje; luego el material citado no se usa como una pieza de lenguaje.

Esta respuesta puede no satisfacer por completo nuestra dudas. Para empezar, están los casos problemáticos en los que es conveniente tanto usar como mencionar la misma expresión diciendo o inscribiendo una única instancia particular de la expresión. Una vez resolví adoptar una forma coherente de emplear la cita en mis escritos profesionales. Mi plan consistía en usar comillas simples cuando quería referir una expresión que contenía una instancia particular de sí misma, y comillas dobles cuando quería usar la expresión con su significado usual indicando al mismo tiempo que la palabra era extraña o especial ("citas de alarma"). Siento vergüenza al admitir que luché con esta fórmula absurda e inmanejable durante un par de años hasta que pude percatarme de que la segunda categoría contenía las semillas de su propia destrucción. Consideremos, por ejemplo, un pasaje anterior de este trabajo en el que digo más o menos:

> Quine dice que la cita "... tiene cierta característica anómala".

¿Las palabras citadas se usan o se mencionan? Obviamente se mencionan, dado que las palabras son de Quine y yo quiero resaltar este

[2] W. V. Quine, *Mathematical Logic*, Cap. 4.
[3] A. Church, *Introduction to Mathematical Logic*, Cap. 8.

hecho. Pero igualmente obvio es el hecho de que las palabras son usadas; si no lo fueran, lo que sigue a la palabra "cita" sería un término singular, y esto no puede ocurrir si he producido una oración gramática. Tampoco resulta fácil modificar mis palabras de manera de resolver la dificultad. Por ejemplo, no basta con escribir "Quine usó las palabras 'tiene cierta característica anómala' para la cita", pues ello deja pendiente lo que él quiso significar con esas palabras. He aquí otro caso mixto de uso y mención que no resulta fácil de discriminar:

> Dhaulagiri está próximo a Anapurna, la montaña cuya conquista Maurice Herzog describió en su libro del mismo nombre.

La última frase "del mismo nombre" no puede significar del mismo nombre que la montaña, pues la montaña tiene muchos nombres. Más bien significa del mismo nombre de la montaña empleada antes en la oración. Llamaría a esto un caso genuino de cita, pues la oración refiere una expresión mediante la exhibición de una instancia particular de dicha expresión; pero es un caso que se las arregla sin comillas.

O consideremos este caso:

> Las reglas de Clouting y de Dragoff son aplicables, en este orden.[4]

Si dejamos momentáneamente de lado estos ultimos ejemplos por considerarlos patológicos y tal vez curables, hay una manera, hoy estándar, de respaldar la idea de que en la cita el material citado no se usa. Es la interpretación de la cita que Tarski propone como la única defendible. Según ella una cita, consistente en una expresión flanqueada por comillas es como una única palabra y debe considerarse lógicamente simple. Las letras y espacios dentro del material citado se ven como accidentes en el deletreo de una palabra más larga y por lo tanto carecen de significado en forma aislada. El nombre de una comilla es así, dice Tarski, como el nombre propio de un hombre.[5] Llamaré a esto la *teoría del nombre propio* de la cita. Church atribuye la misma idea, o al menos un método con las mismas consecuencias, a Frege. Escribe Church:

> Frege introdujo el recurso de indicar sistemáticamente la autonomía mediante comillas, y en sus publicaciones ulteriores (si bien no en *Begriffsschrift*) las palabras y los símbolos usados autónomamente están encerrados entre comillas simples en todos los casos. Esto tiene por efecto que la palabra encerrada entre comillas simples sea tratada como una palabra diferente de la palabra entre comillas —como si las comillas fueran dos letras más en el deletreo de la palabra— y de este modo se elimina el equívoco mediante la provisión de dos palabras diferentes para corresponder a los significados diferentes.[6]

[4] El ejemplo es de J. R. Ross, "Metalinguistic Anaphora", 273.
[5] A. Tarski, "The Concept of Truth in Formalized Languages", 159.
[6] A. Church, *Introduction to Mathematical Logic*, 61-2.

A menos que yo esté en un error, este párrafo exhibe una confusión común. Pues ¿qué expresión es aquélla, de acuerdo con el punto de vista que Church atribuye a Frege, que refiere la palabra de la cual una instancia particular aparece dentro de las comillas? ¿Es ella esa misma palabra (dado el contexto), o la cita como un todo? Church dice que ambas, aunque no pueden ser idénticas. La palabra misma, dado que una expresión entre comillas tiene un significado distinto de su significado usual; está "tratada como una palabra diferente" la cual se usa "autónomamente". La cita como un todo, dado que las comillas son parte del deletreo.

Quine ha promovido repetidamente y en términos muy coloridos la idea de la cita como término singular inestructurado. No sólo niega, como ya hemos visto, que las citas sean descripciones, sino que también afirma que las letras dentro de las comillas en una cita se presentan allí "...como un mero fragmento de un nombre más largo que contiene, además de este fragmento, las dos comillas".[7]

El mérito de esta aproximación a la cita está en el énfasis que pone en el hecho de que la referencia de una cita no puede interpretarse como debida, al menos en alguna forma normal, a la referencia de las expresiones presentes entre las comillas. Pero me parece que, como explicación del funcionamiento de la cita en el lenguaje natural, la aproximación es extremadamente deficiente. Si las citas son términos singulares sin estructura, entonces la *categoría* de los nombres-comilla no encierra más significatividad que la categoría de los nombres que comienzan y terminan con la letra "a" ("Atlanta", "Alabama", "Alta", "Athena", etcétera). Bajo este punto de vista, no hay ninguna relación, más allá de un accidente de deletreo, entre una expresión y el nombre-comilla de esa expresión. Si aceptamos esta teoría, nada se perdería si sustituyéramos cada nombre-comilla por algún nombre no relacionado, pues ésta es la característica de los nombres propios. Y así no queda ningún rastro, por cuanto respecta a esta teoría de la cita, de las reglas informales que gobiernan la cita y que parecen tan claras: que si uno quiere formar un nombre-comilla de una expresión, tiene que flanquear dicha expresión con comillas; y que un nombre-comilla refiere "su interior" (según palabras de Quine). Nada queda de la noción intuitivamente atractiva de que una cita en cierto modo ilustra de qué trata.

Estas objeciones son suficientes en sí mismas para despertar dudas respecto de la afirmación de Tarski de que esta interpretación de la cita es "...la más natural y está en completo acuerdo con la forma acostumbrada de usar las comillas..."[8] Pero hay otra objeción que yo considero decisiva, cual es que en base a esta teoría no podemos dar una explicación satisfactoria de las condiciones bajo las cuales una oración arbitraria que contiene una cita es verdadera. En una teoría adecuada,

[7] W. V. Quine, *From a Logical Point of View*, 140. Compárese con Quine, *Methods of Logic*, 38 y *Word and Object*, 143; también con B. Mates, *Elementary Logic*, 24.

[8] A. Tarski, "The Concept of Truth in Formalized Languages", 160.

se interpreta que cada oración debe su verdad o falsedad a la forma en que se la construyó a partir de una cantidad finita de partes mediante la repetida aplicación de un número finito de modos de combinación. Hay por supuesto, un número finito de nombres-comilla, puesto que cada expresión tiene su propio nombre-comilla y hay un número infinito de expresiones. Pero en la teoría de la cita que estamos considerando, los nombres-comilla no tienen estructura significativa. De allí que una teoría de la verdad no podría abarcar en general a las oraciones que contienen citas. Debemos rechazar la interpretación de la cita como un nombre propio si es que queremos una teoría satisfactoria para un lenguaje que contiene citas.[9]

Me referiré ahora a una teoría de la cita muy diferente, que podría ser llamada *teoría pictórica* de la cita. De acuerdo con este punto de vista, no es la cita toda, esto es, la expresión nombrada más las comillas, lo que refiere la expresión, sino más bien la expresión misma. El papel de las comillas es indicar cómo debemos considerar la expresión en su interior: las comillas constituyen un medio ambiente lingüístico dentro del cual las expresiones hacen algo especial. Este fue tal vez el punto de vista de Reichenbach, quien dijo que las comillas "...transforman un signo en un nombre de ese signo".[10] Quine también sugiere esta idea cuando escribe que una cita "...designa su objeto... ilustrándolo",[11] pues naturalmente sólo del interior de una cita podría decirse que es como la expresión referida (las comillas no están en la ilustración: son su marco). Y Church, en el pasaje antes discutido, también juega con la noción de que en la teoría de Frege "...una palabra encerrada entre comillas simples será tratada como una palabra diferente" en cuanto es usada "autónomamente", es decir, para nombrarse a sí misma.

Debería admitirse que los tres autores recién mencionados vacilan en los pasajes aludidos entre la teoría del nombre propio de la cita y la teoría pictórica. Pero esas teorías son claramente distintas; y si tenemos en mente las deficiencias de la teoría del nombre propio, nos vemos en el deber de considerar a la teoría pictórica por sus propios méritos. A primera vista promete dos ventajas: atribuye *alguna* estructura a las citas, dado que las entiende compuestas por las comillas (que determinan la escena para interpretar su contenido) y por el material citado. E insinúa, en su apelación a la relación de ilustración, una teoría en condiciones de explotar nuestra comprensión intuitiva de la forma en que opera la cita.

Estas ventajas aparentes se esfuman al examinarlas. La dificultad es ésta. Lo que se busca es una explicación de la forma en que la cita nos permite referir expresiones mediante su ilustración. Pero en la presente teoría, las comillas crean un contexto en el cual las expresiones se

[9] Véase el Ensayo 1.

[10] H. Reichenbach, *Elements of Symbolic Logic*, 335. Reichenbach dice otras cosas que apuntan hacia una teoría diferente.

[11] W. V. Quine, *Mathematical Logic*, Cap. 4.

refieren a sí mismas. Entonces ¿cómo se caracteriza la ilustración en esta teoría? Si una expresión dentro de comillas se refiere a sí misma, el hecho de que ella también se ilustre a sí misma es simplememte una irrelevancia colateral.

¿Prestaría alguna ayuda agregar que las comillas crean un contexto en el cual interpretamos su contenido como una ilustración de lo que se refiere? De ningún modo; esto es meramente una manera tendenciosa de decir que la expresión se refiere a sí misma. En breve: una vez que se asigna al contenido de la cita un rol lingüístico normal, el hecho de que pueda parecerse a algo no tiene más significatividad para la semántica que la onomatopeya o que el hecho de que la palabra "polisilábica" sea polisilábica.

Se nos podría escapar otro punto importante. La relación de ilustración entre un objeto y sí mismo es apenas interesante y la teoría, tal como la estamos interpretando, trata vanamente de llegar a algo con esta insulsa idea. Pero la ilustración más interesante que percibimos en la cita no se da entre expresión y expresión. Lo que nos permite en la cita referir cierta expresión, que podemos tomar por una forma abstracta, es el hecho de que tenemos delante de nosotros en la página o en el aire algo que *tiene* esa forma: una instancia particular, escrita o dicha. La teoría pictórica no sugiere una forma de introducir una inscripción o emisión en la ilustración. Esto sólo podría hacerse describiendo, nombrando o señalando la *instancia particular* relevante, y no se ha presentado ningún mecanismo para ese fin.

La teoría pictórica de la cita tiene reminiscencias de la teoría de Frege de los contextos opacos (que nosotros llamamos oblicuos) tales como los creados por "necesariamente", "Jones cree que...", "Galileo dijo que...", etcétera. Hay conspicuas diferencias entre estos contextos tal como los analiza Frege, y la cita tal como la trata la teoría pictórica. En la cita las palabras pueden cambiar su parte del habla (pues cada expresión se vuelve un nombre o descripción) mientras que en los otros contextos esto nunca sucede: y en la cita, mas no en otros contextos opacos, un absurdo tiene sentido. Pero existe la sorprendente similaridad de que en ambos casos se supone que algún recurso lingüístico crea un contexto dentro del cual las palabras adoptan nuevos roles referenciales. Este concepto de un contexto que altera la referencia nunca ha sido explicado convenientemente, y el mismo Frege expresó su recelo: sin duda no se presta el tratamiento directo en una teoría de la verdad. El problema con la teoría pictórica, al igual que con el tratamiento que da Frege a los contextos opacos en general, es que las referencias atribuidas a palabras o expresiones en sus contextos especiales no son funciones de sus referencias en contextos ordinarios, y entonces las expresiones creadoras de contextos especiales (como las comillas o las palabras "dijo que") no pueden ser vistas como expresiones funcionales.[12]

[12] Para una mayor discusión sobre este punto, véase el Ensayo 7.

Un defecto central de la teoría del nombre propio de la cita consiste en que, si bien veía a las citas como expresiones bien formadas del lenguaje, no proporcionaba una teoría articulada que mostrara cómo cada infinitud de tales expresiones debía su referencia a su estructura. El experimento que acabamos de concluir demostró que es posible tratar a las citas como poseedoras de una estructura semánticamente significativa. Avancemos en esta dirección.

Geach ha insistido largamente en que las citas son en realidad *descripciones*, y por lo tanto tienen estructura, y ha criticado a la teoría del nombre propio tal como yo lo he hecho[13] (si bien no relaciona sus críticas con la necesidad de una teoría de la verdad). Su teoría, tal como yo la entiendo, es la siguiente. Una única palabra entre comillas se nombra a sí misma; esto es un nuevo ítem del vocabulario, y no es semánticamente complejo (no estoy seguro de que Geach diga esto último). Hasta ahora, la teoría es igual a la teoría del nombre propio. Pero una expresión más larga que se cita representa una descripción estructurada. Así, " 'Alicia dormía' " abrevia a " 'Alicia' 'dormía' " que se lee "la expresión obtenida de escribir " 'Alicia' " seguido de " 'dormía' ". Esta teoría tiene las ventajas de la teoría fregeana precedente, y es mucho más simple y natural. (Se la podría llamar *teoría del deletreo* de la cita.)

Tanto Tarski como Quine implican, por cosas que dicen, que ven la posibilidad de una teoría similar. De esta manera Tarski afirma que si aceptamos la teoría del nombre, entonces los nombres-comilla pueden eliminarse y reemplazarse en cualquier parte por nombres descriptivos estructurales,[14] mientras que Quine sostiene que podemos disipar la opacidad de la cita, cuando lo deseemos, recurriendo al deletreo.[15] Ambos tienen en mente el mismo recurso que Geach, excepto que Geach toma a las palabras como las unidades más pequeñas, en tanto que Tarski y Quine entienden que éstas son las letras y símbolos individuales. El resultado, en las abreviaturas de la cita ordinaria, es el mismo. En notación primitiva, que revela toda la estructura a simple vista, Geach no encuentra problemas para escribir el lenguaje (pues sólo cada palabra necesita comillas) pero sí tropieza con ellos para aprenderlo o describirlo (tiene un vocabulario primitivo mucho más grande: el doble del tamaño normal si dejamos de lado la iteración).

No hay ninguna dificultad en extender una definición de verdad a los recursos de deletreo que sugieren Quine, Tarski y Geach; pero puede considerarse que estos recursos simplemente están abreviados por la cita usual. Puede respaldarse esta idea de una mera abreviación describiendo un método mecánico que permita utilizar ambos estilos de notación pasando sucesivamente de uno a otro. Así, dado el nombre-comilla " 'Alicia dormía' ", la máquina arranca desde la izquierda reproduciendo

[13] P. T. Geach, *Mental Acts*, 79 y ss. Compárese con Geach, "Quotation and Quantification", 205-9.

[14] A. Tarski, "The Concept of Truth in Formalized Languages", 160.

[15] W. V. Quine, *Word and Object*, 212.

las primeras comillas, luego la letra "A", luego otras comillas, luego un signo de concatenación, luego otras comillas, y así sucesivamente hasta que llega un conjunto de comillas del original. Las reproduce y para. El resultado será:

"A"⌒"l"⌒"i"⌒"c"⌒"i"⌒"a"⌒espacio⌒"d"⌒"o"⌒"r"⌒ "m"⌒"i"⌒"a"

Puesto que las dos notaciones son mecánicamente intercambiables, no hay razón para no considerar que una semántica para la una sea semántica para la otra: luego esto *podría* verse como una teoría de cómo funciona la cita en el castellano (con modificaciones sería aplicable a otros lenguajes). Pero ¿sería ésta una teoría correcta de la cita ordinaria? Hay varias razones para decir que no.

Nótese en primer término que la apariencia de las comillas en la notación expandida resulta extraña. La teoría opera identificando un grupo finito de unidades (palabras o letras) de las que se compone toda expresión del lenguaje a describir. Luego se introducen nombre propios inestructurados de estas unidades, junto con una notación para la concatenación. Una teoría así funciona igualmente bien, y resulta menos equívoca, si se eliminan completamente las comillas y se introducen nuevos nombres de los bloques constructivos. Para ilustrar esto (siguiendo el método de Geach), supongamos que la palabra "alc" nombra a la palabra "Alicia" y que la palabra "dr" nombra a la palabra "dormía"; entonces "Alicia dormía" estaría descrito por:

alc⌒dr

o, si usamos el método de Quine:

A⌒ele⌒i⌒ce⌒i⌒a⌒espacio⌒de⌒o⌒ere⌒eme⌒i⌒a

Este pequeño ejercicio apunta a enfatizar el hecho de que esta teoría no captura nada de la idea de los *signos* de la cita, las comillas; nada de la idea de que uno puede formar el nombre de una expresión arbitraria encerrando a ésta entre comillas. En la teoría del deletreo, ningún ítem articulado del vocabulario corresponde a las comillas, y entonces la teoría no puede reflejar una regla para este uso. La máquina simplemente conoce de memoria el nombre de cada una de las expresiones más pequeñas. Resulta claro que se ha perdido un elemento esencial de la idea de que las citas ilustran acerca de qué tratan.

Una manera llamativa de ver qué es relevante para la estructura y qué no lo es consiste en tratar de aplicar la generalización existencial y la sustitución de identidad. Una forma estándar de demostrar que la cita tal como se la usa normalmente no presenta su estructura en su superficie consiste en observar que de:

"Alicia dormía" es una oración.

no podemos inferir:

($\exists x$) ("x dormía" es una oración).

Tampoco, suponiendo que "alc" nombra a "Alicia", podemos inferir:

"alc dormía" es una oración

ni tampoco:

alc$^\frown$"dormía" es una oración.

Pero (si usamos la versión de Geach de la teoría del deletreo) *podemos* pasar de:

"Alicia dormía" es una oración.

a: "Alicia"$^\frown$"dormía" es una oración.

y de allí a:

alc$^\frown$"dormía" es una oración

y luego a:

($\exists x$) (x $^\frown$"dormía" es una oración)

o:

($\exists x$) (alc$^\frown x$ es una oración).

En la versión de Quine de la teoría, podríamos pasar de:

"Alicia" es una palabra

a:

"A"$^\frown$ "l"$^\frown$"i"$^\frown$"c"$^\frown$"i"$^\frown$"a" es una palabra

a:

($\exists x$) ($\exists y$) ("a" x "i" "y" "i" "a" es una palabra).

Estas derivaciones muestran claramente que las comillas no juegan un papel vital en la teoría del deletreo; y también que esta teoría no es una teoría de cómo opera la cita en el lenguaje natural.

Se ha perdido un elemento esencial de la ilustración, si bien quizá no del todo, pues la teoría del deletreo parece depender de que la des-

cripción de expresiones complejas reproduzca el *orden* de las expresiones descritas. En la descripción que provee la teoría, los nombres de expresiones particulares no necesitan semejar lo que nombran, pero en la descripción como un todo, los nombres de expresiones que están concatenadas están ellos mismos concatenados.

Sin embargo, incluso este residuo de la idea de ilustración es superficial. Las descripciones que proporciona la teoría del deletreo son en sí mismas, desde el punto de vista de un lenguaje totalmente articulado, meras abreviaciones de algo más complicado en lo cual el orden de las expresiones bien puede cambiarse. Pienso que deberíamos concluir que la teoría del deletreo de la cita no tiene conexión con la idea de que entendemos a las citas como expresiones ilustrativas.

Hay otros importantes usos de la cita en un lenguaje natural que no pueden explicarse mediante la teoría del deletreo y que un lenguaje construido en la forma que ella sugiere no podría contener. La teoría del deleteo no podría, al menos bajo ninguna forma obvia, abordar aquellos casos mixtos de uso y mención que discutimos más atrás, ni tampoco, por cierto, algún caso que parezca depender de una referencia demostrativa de una emisión o inscripción. Un uso importante de la cita en el lenguaje natural consiste en introducir nuevas notaciones por medio de su exhibición entre comillas; esto resulta imposible en la teoría del deletreo en caso de que la nueva notación no esté compuesta de elementos que tengan nombres. En la teoría del deletreo tampoco podríamos usar la cita para enseñar un lenguaje extranjero basado en un nuevo alfabeto o notación, como por ejemplo el khmer o el chino. Puesto que éstas son funciones ejecutadas fácilmente por la cita ordinaria (sea con o sin comillas), no podemos aceptar la teoría del deletreo como portadora de una descripción adecuada de la cita en el lenguaje natural.

Hemos descubierto una breve lista de condiciones que una teoría de la cita competente debe satisfacer. La primera de ellas es que, al igual que una teoría para cualquier aspecto de un lenguaje, debería coaligarse con una teoría general de la verdad para las oraciones de lenguaje. Las demás condiciones son específicas para la cita. Una de ellas es que la teoría proporcione una función semántica articulada para los *recursos* de la cita (comillas o equivalentes verbales). Cuando aprendemos a entender la cita aprendemos una regla con innumerables aplicaciones: si uno quiere referir una expresión, puede hacerlo poniendo entre comillas una instancia particular de la expresión que uno quiere mencionar. Una teoría satisfactoria debe incorporar o explicar de algún modo esta aplicación. Y por último, una teoría satisfactoria debe explicar el sentido en el cual una cita ilustra aquello que se refiere, pues de otro modo sería inadecuada para describir importantes usos de la cita, por ejemplo, la introducción de noveles piezas de notación y de nuevos alfabetos.

No es difícil llegar a una teoría satisfactoria una vez que los requerimientos son claros. La dificultad principal surge —tal vez ahora sea obvio— a partir de las exigencias simultáneas de asignar una estructura articulada a las citas y de que éstas ilustren lo que mencionan. Pues

la estructura lingüística articulada debe ser aquí aquella de la descripción, y el describir parece excluir la necesidad de ilustrar. El requisito de una estructura deriva de la exigencia subyacente de una teoría del significado, entendida aquí como una teoría de la verdad; todo lo que se necesita es una estructura suficiente como para implementar la caracterización recursiva de un predicado de verdad. Mas una estructura suficiente todavía será demasiado en tanto consideremos al material citado como parte de la sintaxis semánticamente significativa de una oración. La cura es entonces dejar de lado esta presunción.

Es natural suponer que las palabras que aparecen dentro de los límites de una oración son partes legítimas de la oración; y en el caso de las citas, hemos acordado que las palabras dentro de las comillas nos ayudan a referir dichas palabras. Sin embargo lo que yo propongo es que esas palabras entre comillas no son, desde un punto de vista semántico, parte de la oración. De hecho, ya es confuso decir que ellas son palabras. Lo que aparece entre comillas en una *inscripción*, no una forma, y lo que necesitamos de ella es que ayude a referir su forma. En mi teoría, que podemos llamar *teoría demostrativa* de la cita, la inscripción interior no refiere nada, ni es parte de ninguna expresión que lo haga. Son más bien las comillas las que efectúan todo ese referir, y ayudan a referir una forma señalando algo que la tiene. En la teoría demostrativa, ni la cita como un todo (las comillas más su contenido) ni el contenido sólo son, excepto por accidente, un término singular. El término singular está en las comillas, y se lo podría leer como "la expresión de la cual una instancia particular está aquí". O, para mostrar la forma en que puede ahora decirse genuinamente que la ilustración toma parte: "la expresión con la forma ilustrada aquí".

No desacreditamos esta teoría si decimos que deja de lado el hecho de que el material citado es sintácticamente parte de la oración; si se la toma haciendo abstracción de la semántica, la cuestión de la posición es trivial. En las oraciones habladas, la secuencia temporal ejecuta la función de la disposición lineal en la escritura. Pero si digo "capturé un pez así de grande" o "capturé este pez hoy", mis manos, o el pez, no se tornan parte del lenguaje. Podríamos remover el material citado del corazón de la oración con la suficiente facilidad. La cita es un recurso para señalar inscripciones (o emisiones) y puede usarse, y a menudo se lo hace, para señalar inscripciones o emisiones espacial o temporalmente exteriores a la oración citativa. Así, si tras una afirmación suya yo digo "Nunca se dijo algo tan cierto", estoy refiriendo una expresión, pero lo hago indicando una incorporación de ese algo en una emisión. Las comillas podrían derivarse de manera de extraer el material citado de una oración en la cual aquéllas no desempeñan un papel semántico. De esta manera, en lugar de :

"Alicia dormía" es una oración

podríamos escribir:

Alicia dormía. La expresión de la cual esto es una instancia particular es una oración.

Imaginemos la instancia particular de "esto" reforzada con dedos índices señalando hacia la instancia particular de "Alicia dormía".

Descuento que resulta obvio que la teoría demostrativa asigna a las oraciones que contienen citas una estructura que una teoría de la verdad puede manipular sin complicaciones (naturalmente se supone que hay alguna manera de dar cabida a demostrativos, y a este respecto ya he tratado de indicar que no existe ninguna dificultad real para hacer un lugar a elementos demostrativos o deícticos dentro de una teoría formal de la verdad).[16] Por último, es obvio que se ha explotado y explicado la característica ilustrativa de la cita. Por lo tanto la teoría demostrativa también autoriza el uso de la cita para introducir nuevas partículas de tipografía y para tratar lenguajes con nuevos alfabetos. Para concluir consideraré cómo se las arregla con los casos mixtos de uso y mención que exhibiéramos más atrás.

He dicho que para la teoría demostrativa el material citado no forma parte, semánticamente, de la oración citativa. Pero esto iba más allá de lo necesario o de lo deseable. El recurso de indicar puede usarse sobre todo aquello que esté al alcance del indicador, y no hay razón para que una inscripción en uso activo no se pueda ostentar en el proceso de mención de una expresión. Ya he indicado un tipo de caso importante, y hay muchos más ("Preste usted atención a lo que voy a decir", "¿Por qué usó usted esas palabras?", etcétera). Cualquier instancia particular puede servir de blanco a las flechas de la cita, por lo que en particular una oración citativa puede, después de todo, contener por casualidad una instancia particular con la forma requerida para los propósitos de la cita. Tales instancias particulares cumplen entonces una doble tarea, una como engranajes significativos en la máquina de la oración, otra como objetos semánticamente neutrales con una forma útil. De este modo:

Quine dice que la cita "...tiene cierta característica anómala".

puede transformarse más explícitamente en:

Quine dice, empleando palabras de las cuales éstas son una instancia particular, que la cita tiene cierta característica anómala.

(Aquí el éstas va acompañado por una indicación de la instancia particular de las palabras de Quine.) Para el caso de Anapurna: Dhaulagiri está próximo a Anapurna, la montaña cuya conquista Maurice Herzog describió en su libro con un nombre que tiene esta forma (indicando la instancia particular de "Anapurna").

[16] Véase el Ensayo 2.

Finalmente:

Las reglas de Clouting y de Dragoff son aplicables, en el orden en que estas instancias particulares aparecen (indicando las instancias particulares de "Clouting" y de "Dragoff").

7

Al decir *that**

"Ojalá yo hubiera dicho *that*", dijo Oscar Wilde al alabar una de las agudezas de Whistler. Este, quien tenía una opinión poco favorable respecto de la originalidad de Wilde, replicó: "Lo harás, Oscar, lo harás".[1] Esta anécdota nos recuerda que una expresión como "Whistler dijo *that*" en ocasiones puede servir como una oración gramáticalmente completa. Sugiero que tenemos aquí la clave de un análisis correcto del discurso indirecto, un análisis que abre una vía hacia un análisis de las oraciones psicológicas en general (oraciones acerca de actitudes proposicionales, tal como se las suele llamar), e inclusive, si bien esto está más allá de los alcances del presente trabajo, una pista de lo que distingue a los conceptos psicológicos de otros conceptos.

Pero comencemos con oraciones que usualmente se consideran más representativas del tipo *oratio obliqua*, por ejemplo "Galileo dijo *that* la Tierra se mueve" o "Scott dijo *that* Venus es un planeta inferior". Un problema de estas oraciones es que conocemos su forma lógica. Y admitir esto equivale a admitir que, por más que conozcamos otras cosas acerca de ellas, no conocemos la primera cosa. Si aceptamos que la gramática de superficie es una guía hacia la forma lógica, veremos que "Galileo dijo *that* la Tierra se mueve" contiene la oración "la Tierra se mueve", y que a su vez esta oración consiste en el término singular "la Tierra", y en un predicado, "se mueve". Pero si "la Tierra" es, en este contexto, un término singular, se lo puede reemplazar, en lo que a la verdad o falsedad de la oración continente respecta, por cualquier otro término singular que refiera la misma cosa. No obstante, los que parecen ser reemplazantes apropiados pueden alterar la verdad de la oración original.

La notoria invalidez aparente de esta jugada sólo puede ser aparente, pues la regla en la que se basa no hace sino echar luz sobre lo que envuelve a la idea de un término (lógicamente) singular. Basta, entonces, con unas pocas palabras a modo de explicación: nos hemos equivocado con la forma lógica, o nos hemos equivocado con la referencia del término singular.

* Véase la nota del traductor al pie de página 35 del capítulo 1. [T.]

[1] De H. Jackson, *The Eighteen-Nineties*, 73.

Lo que parece un comportamiento anómalo por parte de los que parecen términos singulares dramatiza el problema de dar una explicación ordenada del discurso indirecto, pero el problema es más agudo. Puesto que lo que toca a los términos singulares toca a lo que ellos tocan, y esto viene a ser todo: cuantificadores, variables, predicados, conectivos. Los términos singulares refieren, o pretenden referir, a las entidades sobre las cuales actúan las variables de cuantificación, y son éstas las entidades respecto de las cuales los predicados son o no son verdaderos. Por lo tanto no debería sorprendernos que, si podemos tropezar con problemas en la oración "Scott dijo *that* Venus es un planeta inferior" si sustituimos "Venus" por "La Estrella del Atardecer", igualmente podamos tropezar con problemas al sustituir "es idéntico a Venus o a Mercurio" en lugar del coextensivo "es un planeta inferior". Las dificultades con el discurso indirecto no se pueden resolver simplemente aboliendo los términos singulares.

¿Qué deberíamos esperar de una consideración adecuada de la forma lógica de una oración? Por encima de todo, diría yo, tal consideración debe llevarnos a ver que el carácter semántico de la oración —su verdad o falsedad— se debe a cómo ella está compuesta: por un número finito de aplicaciones de algunos recursos que forman parte de un número finito de recursos que satisfacen al lenguaje como un todo, a partir de elementos que se extraen de una reserva finita (el vocabulario) que satisface al lenguaje como un todo. Ver una oración bajo esta luz es verla a la luz de una teoría para su lenguaje, una teoría que da la forma de cada oración en ese lenguaje. Una manera de proporcionar una teoría así es mediante la caracterización recursiva de un predicado de verdad, según las pautas que sugirió Tarski.[2]

Dos consideraciones estrechamente relacionadas sostienen la idea de que la estructura que una teoría de la verdad al estilo de Tarski otorga a una oración merece ser llamada forma lógica de la oración. El presentar una teoría así equivale a una demostración persuasiva de que el lenguaje, si bien consiste en un número indefinidamente grande de oraciones, puede ser comprendido por una creatura con poderes finitos. Puede decirse que una teoría de la verdad provee una explicación efectiva del rol semántico de cada expresión significativa para cada vez que ella aparece. Armados de la teoría, siempre podemos responder a la pregunta: "¿Qué están haciendo aquí estas palabras familiares?", diciendo en qué forma constribuyen a la condiciones de verdad de la oración. (Esto no es asignar un "significado", y mucho menos una referencia, a cada expresión significativa.)

A menudo se ve el estudio de la forma lógica de las oraciones a la luz de otro interés, el de dar curso a la inferencia. Desde este punto de vista, dar la forma lógica de una oración es catalogar las características que son de importancia para su localización en la escena lógica, las ca-

[2] A. Tarski, "The Concept of Truth in Formalized Languages". Véanse los Ensayos 2, 4 y 5.

racterísticas que determinan de qué oraciones es ella una consecuencia lógica, y qué oraciones tiene como consecuencias lógicas. Una notación canónica codifica gráficamente la información relevante, lo que vuelve simple a la teoría de inferencia, y mecánica a la práctica donde fuera posible.

Es obvio que ambas aproximaciones a la forma lógica no pueden dar por fruto resultados completamente independientes, pues la consecuencia lógica se define en términos de la verdad. Decir que una segunda oración es una consecuencia lógica de una primera equivale a decir, aproximadamente, que la segunda es verdadera si la primera lo es al margen de cómo se interpretan las constantes no lógicas. Puesto que lo que entendemos por constante lógica puede variar independientemente del grupo de verdades, está claro que las dos versiones de la forma lógica, si bien relacionadas, no necesitan ser idénticas. La relación, en breve, parece ser la siguiente. Toda teoría de la verdad que satisfaga los criterios de Tarski debe tener en cuenta todos los recursos iterativos del lenguaje que afectan la verdad. En los lenguajes familiares para los cuales sabemos cómo definir la verdad los recursos iterativos básicos son reducibles a los conectivos oracionales, al aparato de cuantificación y al operador de descripción si es primitivo. Donde una oración sea una consecuencia lógica de otra sólo en base a la estructura cuantificacional, una teoría de la verdad implicará entonces que si la primera oración es verdadera, la segunda lo es. No tiene sentido, luego, no incluir las expresiones que determinan la estructura cuantificacional entre las constantes lógicas, pues cuando hemos caracterizado la verdad, de la cual depende toda consideración de consecuencia lógica, nos hemos comprometido ya con todo aquello que podría comprometernos al llamar constantes lógicas a tales expresiones. El hacer agregados a esta lista de constantes lógicas incrementará el inventario de verdades lógicas y de relaciones de consecuencia por encima de los requerimientos de una definición de la verdad, y así producirá versiones más ricas de la forma lógica. Para los propósitos del presente trabajo, sin embargo, podemos ajustarnos a las interpretaciones más austeras de la consecuencia lógica y de la forma lógica, aquellas que nos son impuestas cuando enunciamos una teoría de la verdad.[3]

Estamos ahora en condiciones de explicar nuestra aporía del discurso indirecto: ocurre que la relación entre verdad y consecuencia que acabamos de delinear parece desmoronarse. En una oración como "Galileo dijo *that* la Tierra se mueve" el ojo y la mente perciben una estructura familiar en las palabras "la Tierra se mueve". Y allí debe haber una estructura si es que queremos tener una teoría de la verdad, pues un número infinito de oraciones (todas oraciones en el indicativo aparte de algunos problemas sobre el tiempo verbal) producen sentido al colocárselas en el espacio libre en "Galileo diho *that* ____". Luego, si queremos

[3] Para ampliar la defensa de un concepto de forma lógica basado en una teoría de la verdad, véase *Essays on Actions and Events*, 137-46.

dar las condiciones de verdad para todas las oraciones generadas así, no podemos hacerlo oración por oración, sino sólo descubriendo una estructura articulada que nos permita considerar que cada oración está compuesta de un número finito de recursos que efectúan una contribución enunciada a sus condiciones de verdad. Sin embargo, tan pronto como asignamos una estructura familiar debemos dejar que fluyan las consecuencias de tal asignación, y éstas, como sabemos, son consecuencias que para el caso del discurso indirecto nos negamos a aceptar. En cierta forma, la cuestión es aun más extraña que eso. No sólo las consecuencias familiares no fluyen de lo que parece ser una estructura familiar, sino que nuestro sentido común del lenguaje se siente poco seguro de las inferencias basadas en las palabras que siguen al "dijo *that*" del discurso indirecto (hay excepciones).

Entonces, la paradoja es la siguiente: por un lado, la intuición sugiere, y la teoría requiere, que descubramos una estructura semánticamente significativa en las "oraciones-contenido" del discurso indirecto (como llamaré a las oraciones que siguen a "dijo *that*"). Por el otro lado, el fracaso de las relaciones de consecuencia nos invita a tratar a las oraciones contenidas como semánticamente inertes. Pero la forma lógica y las relaciones de consecuencia no pueden estar divorciadas de este modo.

A este respecto, se ha propuesto ver a las palabras que siguen a "dijo *that*" como si operaran entre comillas ocultas, siendo su única función ayudar a referir una oración, y con su inactividad semántica explicada mediante una consideración de cita. Una desventaja de esta propuesta es que ninguna consideración usual de la cita resulta aceptable, ni siquiera dentro de las pautas mínimas que hemos establecido para una consideración de la forma lógica. Puesto que para la mayoría de los autores las citas son términos singulares sin estructura semántica significativa, y dado que debe haber un número infinito de citas diferentes, ningún lenguaje que las contenga puede tener un predicado de verdad definido recursivamente. Esto puede tomarse como demostración de que las consideraciones de la cita que se han obtenido deben estar equivocadas; yo pienso que es así. Pero entonces sería difícil pretender que hemos resuelto el problema del discurso indirecto apelando a la cita.[4]

Tal vez no sea tan trabajoso inventar una teoría de la cita que sirva: la teoría siguiente es bien explícita en Quine. Simplemente véase a las citas como abreviaciones de lo que uno obtiene siguiendo estas instrucciones: a la derecha de la primera letra que tiene comillas iniciales a su izquierda escríbase comillas, luego el signo de concatenación, luego ábrase comillas, en ese orden; hágase esto después de cada letra (tratando a los signos de puntuación como letras) hasta alcanzar el cierre de las comillas finales. Lo que uno tiene ahora es un término singular complejo que da lo que Tarski llama descripción estructural de una expresión. Hay un modesto agregado al vocabulario: nombres de letras y de

[4] Véanse los Ensayos 1 y 6.

signos de puntuación, y el signo de concatenación. Hay un agregado correspondiente a la ontología: letras y signos de puntuación. Y finalmente, si llevamos adelante la aplicación a oraciones en el discurso indirecto, aparecerán las consecuencias lógicas que dicta a nueva estructura. Los dos ejemplos que siguen se vinculan a "Galileo dijo *that* la tierra se mueve":

($\exists x$) (Galileo dijo *that* "la Tie"$^\frown x^\frown x$ $^\frown$"a se mueve")
y con la premisa "R = la 20ª letra del alfabeto"):

Galileo dijo *that* "la tie"$^\frown$la 20ª letra del albafeto$^\frown$la 20ª letra del alfabeto$^\frown$"a se mueve". (He empleado todas las abreviaturas que he podido.) Estas inferencias no representan en sí mismas una crítica a la teoría de la cita; simplemente la iluminan.

Quine discute la aproximación citativa al discurso indirecto en *Word and Object*,[5] y la abandona por lo que, a mi juicio, es una razón errónea. No es que no haya una buena razón; pero *apreciarla* es estar a un paso de una solución, como traté de explicar.

Sigamos los pasos que llevaron a Quine a rechazar la aproximación citativa. La versión de la teoría que él considera no es la misma que propusiera una vez Carnap en el sentido de que "dijo *that*" es un predicado diádico que vale para pares ordenados de gente y de oraciones.[6] El inconveniente de esta idea no es que nos fuerce a asimilar el discurso indirecto al directo, pues no lo hace. El "dijo *that*" del discurso indirecto, como el "dije" del directo, puede relacionar personas y oraciones, pero ser una relación diferente; lo primero, a diferencia de lo último, puede ser veradadero para una persona, y para una oración que nunca dijo en un lenguaje que nunca conoció. El problema radica más bien en la posibilidad de que la misma oración pueda tener diferentes significados en diferentes lenguajes, una posibilidad no muy lejana por cierto si contamos a los idiolectos como lenguajes. Para dar un ejemplo, los sonidos *"Empedokles liebt"* funcionan bien como una oración del alemán o del inglés, en un caso diciendo que Empédocles amaba y en el otro contándonos lo que hacía en la cima del Etna. Si analizamos "Galileo dijo *that* la Tierra se mueve" y concluimos que afirma una relación entre Galileo y la oración "La Tierra se mueve", no podemos presumir que Galileo hablaba inglés, pero no podemos evitar la presunción de que las palabras de la oración-contenido deben entenderse como una oración del inglés.[7]

Decir que la relatividad al inglés es una presunción puede llevar a confusión; quizá la referencia al inglés sea explícita, como veremos. Una versión expandida de nuestra oración favorita podría ser "Galileo

[5] W. V. Quine, *Word and Object*, Cap. 6. En adelante los números entre paréntesis refieren páginas de este libro.

[6] R. Carnap, *The Logical Syntax of Languages*, 248. Lo mismo fue propuesto en efecto por P. T. Geach en *Mental Acts*.

[7] Este punto lo debemos a A. Church, "On Carnap's Analysis of Statements of Assertion and Belief".

dijo una oración que significaba en su lenguaje lo que 'La Tierra se mueve' significa en inglés". Dado que en esta versión son necesarias todas las palabras excepto "Galileo" y "La Tierra se mueve" para hacer el trabajo de "dijo *that*", debemos considerar la referencia al inglés explícita en "dijo *that*". Para ver cuán singular es esto, sin embargo, sólo hace falta reflejar que las palabras inglesas "*said that*", con su referencia interna al inglés, no traducirían ya (ni siquiera para los más básicos requisitos extensionales) las francesas "*dit* que".

Podemos desviar la dificutad de traducción del "*said that*" o "*dit* que" si consideramos a estas expresiones como predicados triádicos que relacionan un hablante, una oración y un lenguaje, con la referencia a un lenguaje provista tanto por nuestro (en la práctica casi infalible) conocimiento del lenguaje al cual se considerará que pertenece el material citado, o por una referencia demostrativa al lenguaje de la oración completa. Cada una de estas sugerencias ejerce su propia atracción, pero ninguna lleva a un análisis que pueda pasar el test de traducción. Si tomamos la propuesta demostrativa, la traducción al francés convertirá "*said that*" en "*dit* que", y la referencia demostrativa pasará automáticamente, y por tanto quizás aun dentro de los límites de la traducción estricta, del inglés al francés. Pero cuando traducimos el término singular final, que nombra una oración inglesa, obtenemos un resultado palpablemente falso.

Estos ejercicios ayudan a sacar a la luz importantes características de la aproximación citativa. Pero ahora es tiempo de señalar que existiría una anomalía en una posición que, como la que está en consideración, abjurara de la referencia a proposiciones en favor de la referencia a lenguajes. Y esto porque los lenguajes (como lo señala Quine en un contexto similar en *Word and Object*) están por lo menos tan mal individuados como las proposiciones, mayormente por las mismas razones. Por cierto una propuesta obvia que los vincula es la siguiente: los lenguajes son idénticos cuando oraciones idénticas expresan proposiciones idénticas. Vemos, entonces, que las teorías que aplican la cita al discurso indirecto, como las que hemos visto, no pueden proclamar una ventaja sobre las teorías que introducen con franqueza entidades intensionales desde un principio; consideremos brevemente entonces teorías de este último tipo.

Podría pensarse, y tal vez a menudo se lo hace, que si deseamos dar la bienvenida a las entidades intensionales sin restricciones —propiedades, proposiciones, conceptos individuales, y cualesquiera otras— luego no hay más dificultades para describir la forma lógica de oraciones en *oratio obliqua*. Esto no es así. Ni los lenguajes que Frege sugiere como modelos para los lenguajes naturales ni los lenguajes que describe Church están sujetos a la teoría en el sentido de una definición de la verdad que responda a los requisitos de Tarski.[8] Lo que se interpone en

[8] G. Frege, "On Sense and Reference"; A. Church, "A Formulation of the Logic of Sense and Denotation". Véase el Ensayo 1.

el camino para el caso de Frege es que cada expresión referente tiene un número infinito de entidades que puede referir, dependiendo del contexto, y no hay ninguna regla que dé la referencia en contextos más complejos sobre la base de la referencia en otros más simples. En los lenguajes de Church, hay un número infinito de expresiones primitivas; esto entorpece directamente la posibilidad de caracterizar recursivamente un predicado de verdad que satisfaga los requerimentos de Tarski.

Las cosas podrían emparcharse si se sigue una idea central de *Meaning and Necessity,* de Carnap, y se limita los niveles semánticos a dos: extensiones e intensiones (de primer nivel).[9] Así, una estrategia atractiva podría ser dar vuelta a Frege, simplificado de ese modo, dejando que cada término singular refiera su sentido o intensión, y proveyendo una función de realidad (similar a la función delta de Church) para relacionar intensiones con extensiones. Bajo este tratamiento nuestra oración de prueba emergería de esta manera: "La realidad de Galileo dijo *that* la Tierra se mueve". Aquí debemos suponer que "la Tierra" nombra un concepto individual al cual la función referida por "se mueve" relaciona con la proposición de que la Tierra se mueve; la función referida por "dijo *that*" por su parte relaciona Galileo y la proposición de que la tierra se mueve con un valor de verdad. Por último, el nombre "Galileo" refiere un concepto individual que está relacionado por la función referida por "la realidad de" con Galileo. Si se aplica ingenio, tal vez pueda lograrse que esta teoría haga un lugar a cuantificadores que liguen variables tanto dentro como fuera de los contextos que crean verbos como "dijo" y "cree". No hay ningún problema especial para definir la verdad para un lenguaje así: todo demuestra ser puramente extensional salvo en su ontología. Parece que esta teoría podría hacer todo lo que hemos pedido. Escrúpulos nominalistas al margen, ¿por qué no aceptarla?

Mis razones en contra de este camino son esencialmente las mismas de Quine. Encontrar las palabras correctas de mi propia cosecha para comunicar lo dicho por otro representa un problema para la traducción (216-17). Las palabras que uso en el caso particular pueden verse como productos de mi teoría total (por vaga y sujeta a corrección que sea) de lo que el hablante originario significa con cualquier cosa que dice: una teoría así es indistinguible de una caracterización de un predicado de verdad, con su lenguaje como lenguaje objeto y con el mío como metalenguaje. El punto crucial es que habrá teorías alternativas igualmente aceptables que diferirán en la asignación de oraciones claramente no sinónimas de la mía como traducciones de la misma emisión del hablante originario. Esta es la tesis de Quine de la indeterminación de la traducción (218-21).[10] Un ejemplo ayudará a develar el hecho de que

[9] La idea de una aproximación esencialmente fregeana limitada a dos niveles semánticos ha sido sugerida por M. Dummett en *Frege: Philosophy of Language*, Cap. 9.

[10] Mi asimilación de un manual de traducción a una teoría de la verdad no aparece en Quine. Para más detalles sobre esto y cuestiones relacionadas, véanse los Ensayos 2, 11 y 16.

esta tesis se aplica no sólo a la traducción entre hablantes de lenguajes conspicuamente diferentes, sino también a casos más próximos a nosotros.

Supongamos que alguien dice (y ahora el discurso es directo): "Hay un hipopótamo en la nevera", ¿tengo necesariamente razón si anuncio que él ha dicho que hay un hipopótamo en la nevera? Tal vez; pero ante nuestras preguntas él continúa. "Es redondeado, tiene la piel arrugada, no le molesta que lo toquen. Tiene un sabor agradable, al menos su jugo, y cuesta unos pocos centavos. Exprimo dos o tres para el desayuno." Tras una cantidad finita de tal conversación transpasamos la línea hasta la cual es admisible o incluso posible decir correctamente que él dijo que había un hipopótamo en la nevera, pues queda claro que él significa algo distinto, al menos con algunas de sus palabras, de lo que yo significo. La hipótesis más simple llegados a ese punto es que mi palabra "hipopótamo" ya no traduce su palabra "hipopótamo"; mi palabra "naranja" lo haría mejor. Pero en cualquier caso, mucho antes de alcanzar el punto donde la traducción homofónica debe abandonarse, la caridad invita a desviaciones. La hesitación respecto de traducir una emisión de otro por una u otra de varias oraciones no sinónimas mías no necesariamente refleja una falta de información: es sólo que más allá de cierto punto no se puede decidir, ni siquiera en principio, entre la idea de que el Otro ha usado las palabras tal como las usamos nosotros pero tiene creencias más o menos extrañas, y la idea de que lo hemos traducido mal. Acorralados entre la necesidad de encontrar el sentido de las palabras de un hablante y la necesidad de encontrar el sentido de su patrón de creencias, lo mejor que podemos hacer es elegir una teoría de la traducción que maximice el acuerdo. Seguramente no tiene futuro suponer que al emitir con toda seriedad las palabras "Hay un hipopótamo en la nevera" el Otro ha discrepado con nosotros acerca de lo que puede haber en la nevera si también nosotros debemos vernos luego discrepando con él acerca del tamaño, forma, color, fabricante, caballos de fuerza y consumo de combustible de los hipopótamos.

Nada de esto demuestra que no cabe la posibilidad de relatar correctamente, mediante el discurso indirecto, lo que otro ha dicho. Todo lo que la indeterminación demuestra es que si hay una manera de hacerlo correctamente, hay también otras maneras que difieren sustancialmente en que se usan oraciones no sinónimas después de "dijo *that*". Y esto es suficiente para justificar nuestra impresión de que hay algo espurio en torno del rigor que las cuestiones de significado deben tener en principio si los significados son entidades.

Esa lección estaba implícita en una discusión iniciada algunos años atrás por Benson Mates. Mates sostenía que la oración "Nadie duda de que quienquiera que crea que el séptimo consulado de Mario duró menos de una quincena cree que el séptimo consulado de Mario duró menos de una quincena" es verdadera pero bien podría volverse falsa si la última palabra fuera reemplazada por las palabras (supuestamente sinónimas) "período de catorce días", y que esto podría ocurrir independientemente de cuáles normas de sinonimia adoptáramos sin

llegar a un preestablecido "sustituible en todas partes *salva veritate*".[11] Church y Sellars respondieron diciendo que la dificultad podía resolverse si se distinguía con firmeza entre sustituciones basadas en el uso que el hablante da al lenguaje y sustituciones teñidas por el uso atribuido a otros.[12] Pero esto es una solución sólo si pensamos que hay alguna manera de diferenciar, en lo que otro dice, qué responde a los significados que él da a sus palabras y qué a sus creencias acerca del mundo. Según Quine, esta distinción no se puede realizar.

El rodeo ha sido largo; vuelvo ahora al tratamiento que da Quine a la aproximación citativa en *World and Object*. Como dijéramos antes, Quine rechaza la relativización a un lenguaje con el argumento de que el principio de individuación de los lenguajes es oscuro, y el beneficio irrelevante para el discurso indirecto cuando los lenguajes son idénticos. (214). El autor sugiere ahora que en lugar de interpretar que la oración-contenido del discurso indirecto tiene lugar en un lenguaje la interpretemos como dicha por un hablante en un tiempo. El hablante (y el tiempo) en relación con el cual la oración-contenido necesita comprensión es, desde luego, el hablante de esa oración, quien de ese modo está atribuyendo indirectamente un dicho a otro. Así, "Galileo dijo *that* la Tierra se mueve" viene ahora a significar algo así como "Galileo habló una oración que en sus labios significaba o que 'La Tierra se mueve' significaba ahora en los míos". Quine no hace objeción a esta propuesta pues cree que tiene reservado algo más simple y por lo menos igualmente bueno. Pero en mi opinión la presente propuesta merece una consideración más seria, pues pienso que es casi correcta, en tanto que las alternativas preferidas por Quine padecen de serios defectos.

La primera de estas alternativas es la teoría inscripcional de Scheffler.[13] Este sugiere que las oraciones del discurso indirecto relacionan un hablante y una emisión: el rol de la oración-contenido consiste en ayudar a transmitir de qué tipo de emisión se trataba. Lo que obtenemos de esta manera es, "Galileo habló una emisión *that* -la-Tierra-se-mueve" el predicado "*x*" es-una-emisión-*that*-la-Tierra-se-mueve" tiene, por cuanto a las teorías de la verdad y de la inferencia respecta, la forma de un predicado inestructurado monádico, Quine no presenta la cuestión de esta manera, y podría resistirse a mi apropiación de los términos "forma lógica" y "estructura" para propósitos que excluyen la aplicación al predicado de Scheffler. Quine llama "compuesto" al predicado y al describirlo dice que se compone de un operador y de una oración (214, 215). Estas son cuestiones de terminología; la sustancia, respecto de la cual no tendría que haber discrepancias, es que en la teoría de Scheffler las oraciones en *oratio obliqua* no tienen relaciones lógicas que dependan de una estructura dentro del predicado, y un predicado

[11] B. Mates, "Synonymity". El ejemplo es de Church.
[12] A. Church, "Intensional Isomorphism and Identity of Belief"; W. Sellars, "Putnam on Synonymity and Belief".
[13] I. Scheffler, "An Inscriptional Approach to Indirect Quotation".

de verdad que sea aplicable a todas esas oraciones no puede ser caracterizado según el estilo de Tarski. La razón es sencilla: hay un número infinito de predicados con la sintaxis "*x* es-una-emisión- —" cada uno de los cuales no está relacionado con el resto según el punto de vista de la teoría semántica.

Quine ha tomado una de las puntas del dilema. Dado que atribuir estructura semántica a las oraciones-contenido del discurso indirecto aparentemente nos fuerza a establecer relaciones lógicas que no queremos, Quine renuncia a la estructura. Esto da por resultado el abandono de otro *desideratum* de la teoría, definir la verdad.

Consistente con su política de renunciar a la estructura que no sostenga inferencias que valga la pena conservar, Quine contempla un paso adicional; escribe: "...una alternativa final que encuentro tan atractiva como cualquier otra es simplemente prescindir de los objetos de las actitudes proposicionales" (216). Donde Scheffler aún veía "dijo *that*" como un predicado diádico que relaciona hablantes y emisiones, pero unificando las oraciones-contenido en predicados monádicos de una sola pieza válidos para las emisiones, Quine ahora concibe la oración-contenido y "dijo-*that*" unificados directamente para formar el predicado monádico "*x* dijo-*that*-la-Tierra-se-mueve", válido para personas. Por supuesto que algunas inferencias inherentes del esquema de Scheffler ahora caducan: ya no podemos inferir "Galileo dijo algo" a partir de nuestra oración de prueba, ni podemos inferir de ella y de "Alguien negó *that* la Tierra se mueve" la oración "alguien negó lo que Galileo dijo". Pero como Quine nos hace notar, inferencias como éstas pueden fallar también en el análisis de Scheffler cuando se extiende el análisis según una línea obvia hasta la creencia y otras actitudes proposicionales, puesto que las emisiones requeridas pueden no llegar a materializarse. (215) Las ventajas de la teoría de Scheffler sobre la "alternativa final" de Quine son así escasas e inciertas; es por esta razón que Quine concluye que la propuesta que invita a la menor cantidad de inferencias es "tan atractiva como cualquier otra".

Desafortunadamente, esta manera de eliminar inferencias indeseadas elimina casi toda la estructura requerida por la teoría de la verdad. Por ello vale la pena volver a dar otra mirada a la propuesta previa de analizar el discurso indirecto en términos de un predicado que relaciona un hablante original, una oración, y el presente hablante de una oración en el discurso indirecto. Pues esa propuesta no interferiría con ninguna de las simples implicaciones que hemos estado discutiendo, y prometía, exclusivamente para las sugerencias recientes y asociada a una teoría viable de la cita, rendir frutos para los métodos semánticos usuales. Pero existe una sutil falla.

Tratamos ya de extraer la esencia del análisis al que hemos vuelto reescribiendo nuestra oración favorita como "Galileo emitió una oración que significaba en sus labios lo que 'La Tierra se mueve' significa ahora en los míos". No debemos pensar mal de esta verborrágica versión de "Galileo dijo *that* la Tierra se mueve" en razón de una aparente referencia a un significado ("lo que 'La tierra se mueve' significa"); esta expre-

sión no será tratada en la teoría como un término singular. Ciertamente se nos pide comprender un juicio de sinonimia entre emisiones, pero no como el fundamento de una teoría del lenguaje, sino meramente como una parte inanalizada del contenido de la forma idiomática familiar del discurso indirecto. La idea que subyace en nuestra torpe paráfrasis es la de *igual-decir*: cuando yo digo que Galileo dijo *that* la Tierra se mueve, represento a Galileo y a mí mismo como igual-dicentes.[14]

Ahora bien, la falla es ésta. Si yo meramente *digo* que Galileo y yo somos igual-dicentes, todavía tengo que *hacer* que lo seamos; ¿y cómo puedo llegar a esto? Obviamente, diciendo lo que él dijo; no usando sus palabras (necesariamente), sino usando palabras del mismo valor aquí y ahora que las de él allá y entonces. Pero esto es precisamente lo que no puedo hacer en la teoría. Pues la teoría presenta a la oración-contenido encerrada entre comillas, y en cualquier teoría estándar de la cita esto significa que la oración-contenido es mencionada y no usada. Al emitir las palabras "La Tierra se mueve" yo, de acuerdo con esta idea, no digo nada remotamente parecido a lo que se atribuye a Galileo haber dicho; de hecho, no digo nada. Mis palabras dentro del marco provisto por "Galileo dijo *that* ——" simplemente ayudan a referir una oración. Sin riesgo de cometer un error, podemos expandir la cita en la forma que lo hicimos poco más atrás. Toda intimación en el sentido de que Galileo y yo somos igual-dicentes se esfuma en esta versión:

Galileo dijo *that* "L"∩ 'a" ∩" "∩ "t" ∩ "i"∩ "e"∩ "r" ∩"r" ∩ "a" ∩ " " ∩"s" ∩ "e"∩" "∩ "m"∩ "u" ∩ "e" ∩ "v"∩ "e".

Parece que hemos sido arrastrados por un accidente notacional, una forma de referir expresiones que cuando se la abrevia produce ilustraciones enmarcadas de las palabras mismas que se ha referido. La dificultad es extraña; veamos si podemos sorterla. Imaginemos un caso alterado. Galileo emite sus palabras "Eppur si muove", yo emito mis palabras, "La Tierra se mueve". Aún no encontramos problemas para reconocer que somos igual-dicentes; una emisión mía coincide con una emisión suya en cuanto dan a entender lo mismo. No estoy usando ahora mis palabras para ayudar a referir una oración; hablo por mí mismo, y mis palabras refieren en la forma usual a la tierra y su movimiento. Si la emisión de Galielo *"Eppur si muove"* nos hizo igual-dicentes, luego

[14] Si hablamos estrictamente, el verbo "dijo" está analizado aquí como un predicado triádico que vale para un hablante (Galileo), una emisión del hablante (*"Eppor si muove"*) y una emisión del atribuidor ("La Tierra se mueve"). Este predicado es, desde un punto de vista semántico, un primitivo. El hecho de que una paráfrasis informal del predicado apele a una relación de igualdad de contenido como ocurre entre emisiones no introduce entidades o semántica intensionales. Algunos han visto esto como una especie de trampa, pero la política es deliberada y se apoya en principios. Para una discusión de la distinción entre cuestiones de forma lógica (que es lo que ahora nos ocupa) y el análisis de predicados individuales, véase el Ensayo 2. También vale la pena observar que la interpretación radical, si tiene éxito, proporciona un concepto adecuado de sinonimia como el que se da entre emisiones. Véase el final del Ensayo 12. (Nota agregada en 1982.)

una u otra emisión de Galileo nos hizo igual-dicentes. La forma "($\exists x$) (la emisión x de Galileo y mi emisión y nos hace igual-dicentes)" es así una manera de atribuir a Galileo cualquier dicho que a mí me plazca en tanto encuentre una manera de reemplazar "y" por una palabra o frase que refiera una apropiada emisión mía. E indudablemente hay una manera en que yo puedo hacer esto: sólo necesito presentar la emisión requerida y reemplazar "y" por una referencia a ella. Helo aquí:

La Tierra se mueve.

($\exists x$) (la emisión x de Galileo y mi última emisión nos hacemos igual-dicentes).

Basta con una abreviatura definicional para reducir este pequeño esquema a:

La Tierra se mueve.
Galileo dijo *that*.

Aquí el "*that*" es un término demostrativo singular que refiere una emisión (no una oración).

Esta forma tiene la pequeña desventaja de que deja al oyente en el aire con respecto al propósito que se persigue al decir "La Tierra se mueve" hasta que el acto ha sido ejecutado. Es como si, digamos, yo contara primero una historia y luego agregara: "Así era en una época". Esto le da cierto toque gracioso, y en todo caso ningún anuncio acerca de cuál es la fuerza ilocutoria de nuestras emisiones podrá asegurar que ellas tengan esa fuerza. Pero en el caso en cuestión nada nos impide revertir el orden de cosas; así:

Galileo dijo *that*.
La Tierra se mueve.

Ahora se puede admitir un minúsculo cambio ortográfico, un cambio sin significatividad semántica, pero que sugiere al ojo la relación de introductor e introducido: podemos suprimir el punto después de "*that*" y la consecuente capitalización es:

Galileo dijo *that* la Tierra se mueve.

Quizá no deba causar sorpresa el notar que la forma de las oraciones psicológicas del inglés aparentemente evolucionó en gran medida según la forma que estas reflexiones sugieren. Según el *Oxford English Dictionary*:

> Se considera por lo general que el uso de *that* emergió del pronombre demostrativo que señala la cláusula a la cual introduce. Cotéjese: (1) "He once lived here: we all know *that*" (En una época vivió aquí: todos sabemos

eso); (2) "*That* (now *this*) we all know: he once lived here" [*Eso* (ahora *eso*) lo sabemos todos: en una época vivió aquí]; (3) "We all know *that* (or *this*): he once lived here" [Todos sabemos *eso* (o *esto*): en una época vivió aquí); (4) "We all know *that* he once lived here..." (Todos sabemos *que* en una época vivió aquí...).[15]

La propuesta entonces es la siguiente: las oraciones del discurso indirecto, tal como sucede, llevan su forma lógica a la vista (la única excepción es un pequeño punto). Estas oraciones consisten en una expresión que refiere a un hablante, el predicado diádico "dijo", y un demostrativo que refiere una emisión. Período. Lo que sigue da el contenido del dicho del sujeto, pero no tiene una conexión lógica o semántica con la atribución original de un dicho. Esto último constituye sin duda la novedad, y de ello depende todo: desde un punto de vista semántico la oración-contenido del discurso indirecto no está contenida en la oración cuya verdad cuenta, es decir la oración que termina con "*that*".

Al tratar con este sujeto, nos convendría hablar de inscripciones y emisiones y actos de habla, y evitar la referencia a oraciones.[16] Y esto porque lo que hace una emisión de "Galileo dijo *that*" es anunciar otra emisión. Como toda emisión, esta primera puede ser seria o tonta, asertiva o traviesa; pero si es verdadera, debe estar seguida de una emisión sinónima de alguna otra. La segunda emisión, el acto presentado, también puede ser verdadera o falsa, hecha a modo de aserción o de travesura. Pero si es tal como se la anunció, al menos debe cumplir con el propósito de conducir el contenido de lo que alguien dijo. El rol de la emisión introductoria no nos es desconocido: nosotros hacemos lo mismo con palabras como "Esto es una broma", "Esto es una orden", "El exigió eso", "Ahora escucha esto". Tales expresiones podrían ser llamadas realizativos, pues se las emplea para anunciar realizaciones por parte del hablante. Sobreviene cierto efecto reflexivo de interés cuando los realizativos se dan en la primera persona del tiempo presente, pues luego el hablante emite palabras que de ser verdaderas lo son exclusivamente por el contenido y modo de la realización que sigue, y el modo de esta realización bien puede estar determinado en parte por la misma presentación realizativa. He aquí un ejemplo que también dará ocasión para un comentario final sobre el discurso indirecto.

"Jones aseveró *that* Entebbe es ecuatorial", vendrá a significar, si aplicamos el análisis del discurso indirecto, algo así como "Una emisión de Jones en el modo asertivo tenía el contenido de esta emisión mía. En-

[15] J. A. H. Murray *et al* (comps.), *The Oxford English Dictionary*, 253. Cotéjese C. T. Onions, *An Advanced English Syntax*, 154-6. Leí por primera vez que "*that*" en tales contextos evolucionó a partir de un demostrativo explícito en J. Hintikka, *Knowledge and Belief*, 13. Hintikka señala que una evolución similar ha tenido lugar en el alemán y en el finlandés. Debo la referencia del *OED* a Eric Stiezel.

[16] Doy por sentado que una teoría de la verdad para un lenguaje que contiene demostrativos debe aplicarse estrictamente a emisiones y no a oraciones, pues si no tratará a la verdad como una relación entre oraciones, hablantes y tiempos. Véanse los Ensayos 2 y 4.

tebbe es ecuatorial". El análisis no sucumbe pues los modos de emisión de los dos hablantes puede diferir; todo lo que la verdad del realizativo requiere es que la segunda emisión, cualquiera sea su modo (asertivo o no), sea igual en contenido a una emisión asertiva de Jones. El que tal asimetría sea apropiada en el discurso indirecto depende de qué grado de aserción leamos dentro del concepto de decir. Ahora supongamos que digo: "Yo asevero *that* Entebbe es ecuatorial". Por supuesto que al decir esto puedo no aseverar nada; el modo de las palabras no puede garantizar el modo de emisión. Pero si mi emisión del realizativo es verdadera, entonces sí estoy diciendo algo en el modo asertivo que tiene el contenido de mi segunda emisión; esto es, si asevero que Entebbe es ecuatorial. Si realmente asevero esto, un elemento de mi éxito es sin duda mi emisión del realizativo, el cual anuncia una aserción. De esta manera los realizativos tienden a ser autosatisfacientes. Tal vez sea este rasgo de los realizativos lo que ha inducido erróneamente a algunos filósofos a pensar que los ejecutores, o sus emisiones, no son ni verdaderos ni falsos.

En el análisis del discurso indirecto que aquí se propone, los problemas habituales parecen encontrar una solución justa. El aparente fracaso de las leyes de sustitución extensional se explica como consecuencia de que consideremos erróneamente lo que en realidad son dos oraciones como una sola: hacemos sustituciones en una oración, pero es la otra (su emisión) la que varía en su verdad. Puesto que una emisión de "Galileo dijo *that*" y cualquier emisión que le siga son semánticamente independientes, no hay razón para predecir, sólo en base a la forma, ningún efecto *particular* sobre la verdad de la primera a partir de un cambio en la segunda. Por otra parte, si la segunda emisión hubiera sido diferente de algún modo, la primera emisión *podría* haber tenido un valor de verdad diferente, pues la referencia del "*that*" habría cambiado.

La paradoja, el que las oraciones (emisiones) *en oratio obliqua* no tengan las consecuencias lógicas que deberían tener si es que la verdad va a ser definida, está resuelta. Lo que sigue al verbo "dijo" tiene solamente la estructura de un término singular, usualmente el demostrativo "*that*". Suponiendo que el "*that*" refiere, podemos inferir que Galileo dijo algo a partir de "Galileo dijo *that*"; pero esto es bienvenido. Las palabras familiares que acompañan al realizativo del discurso indirecto poseen, sí, una estructura, según mi parecer, pero se trata de una estructura familiar y no ocasiona problema alguno para la teoría de la verdad que no estuviera ya presente antes de que se tocara el tema del discurso indirecto.

Desde Frege, los filósofos han debido habituarse a la idea de que las oraciones-contenido del habla acerca de actitudes proposicionales pueden extrañamente referir entidades tales como intensiones, proposiciones, oraciones, emisiones e inscripciones. Lo extraño no son las entidades, que están correctamente en su espacio (si lo tienen), sino la noción de que palabras comunes para planetas, gente, mesas e hipopótamos puedan en el discurso indirecto abandonar estas referencias pedestres a cambio de lo exótico. Si pudiéramos recuperar nuestra inocen-

cia semántica pre-fregeana, pienso que nos parecería sencillamente increíble que las palabras "La Tierra se mueve", emitidas después de las palabras "Galileo dijo *that*", signifiquen algo diferente, o refieran alguna otra cosa que lo que acostumbran cuando aparecen en otros entornos. No cabe duda de que su rol en *oratio obliqua* es en algún sentido especial; pero eso es otra historia. El lenguaje es el instrumento que es puesto que la misma expresión, sin variar sis características semánticas (su significado), puede servir a incontables propósitos. He tratado de mostrar que nuestra comprensión del discurso indirecto no afecta esta concepción básica.

8

Modos y ejecuciones

Frege sostuvo que una explicación adecuada del lenguaje requiere que atendamos a tres características de las oraciones: referencia, sentido y fuerza. En toda oportunidad he insistido en que una teoría de la verdad diseñada en base a una definición de verdad del tipo de Tarski nos dice todo lo que necesitamos saber acerca del sentido.[1] Si contamos a la verdad dentro del dominio de la referencia, como lo hizo Frege, el estudio del sentido se reduce de este modo al estudio de la referencia.

Pero ¿qué hay de la fuerza? En este trabajo pretendo considerar a la fuerza en la única forma bajo la cual estoy seguro de que ella es un rasgo de las oraciones, esto es, en tanto sirve para distinguir los modos. La pregunta que me interesa es, ¿puede una teoría de la verdad explicar las diferencias entre los modos?

Al tratar de responder a esta pregunta estoy respondiendo, por desgracia demasiado tarde,[2] a un desafío al que me incitara Yehoshua Bar-Hillel algunos años atrás; me preguntó cómo sería posible representar el modo dentro de los confines de una teoría de la verdad.

El análisis del modo es interesante, entre otras razones, porque nos obliga a prestar atención a las relaciones entre lo que las oraciones significan y sus usos. Tenemos por un lado la distinción sintáctica, y presumiblemente semántica, entre modos (como: indicativo, imperativo, optativo, interrogativo), y por el otro lado la distinción entre usos de las oraciones (como: hacer aserciones, dar órdenes, expresar deseos, hacer preguntas).

Los modos clasifican oraciones, mientras que los usos clasifican emisiones; pero los modos clasifican indirectamente a las emisiones, pues todo aquello que distingue a las oraciones puede emplearse para ditinguir a las emisiones de ellas. Entonces podemos preguntar, ¿cuál es la relación entre estas dos formas de clasificar emisiones; cómo se relacionan las aserciones con las emisiones de oraciones indicativas, por ejemplo, o las órdenes con las emisiones de oraciones imperativas?

[1] Véanse los Ensayos 2, 4 y 9-12 para la defensa de esta posición, y las dudas que genera.

[2] La conferencia en la cual se leyó este trabajo había sido convocada para celebrar el sexagésimo cumpleaños de Bar-Hillel. Una gran pérdida para la filosofía y para sus muchos amigos, falleció antes de que la conferencia tuviera lugar.

La sugerencia más simple sería que las clases asociadas de emisiones son idénticas: las emisiones de imperativas son órdenes, las emisiones de interrogativas son preguntas, etcétera. Esta idea parece encontrar sustento en parte del trabajo de Dummett. Es así como Dummett explica el uso que da Frege al signo de aserción o rasgo de juicio:

> El rasgo de juicio es propiamente el signo de aserción, aquel que lleva la fuerza asertiva. Por ello no es una expresión funcional, o parte de una: no podemos inquirir cuál es su sentido, o cuál es su referencia; contribuye al significado del símbolo oracional complejo de un modo muy diferente... sólo de la oración a la que se prefija el rasgo de juicio puede decirse que expresa un sentido o que sostiene un valor de verdad: la expresión toda con el rasgo de juicio no expresa nada ni sostiene nada, *afirma* algo; esto es, afirma que el pensamiento expresado por lo que sigue al rasgo de juicio es verdadero.[3]

Aquí Dummett dice que son las oraciones las que hacen aserciones, en tanto que yo pienso que sería más natural decir que una aserción es una emisión, y es el hablante quien hace la aserción. No obstante, esto puede no ser más que un mal terminológico; lo que me preocupa es la pretensión implícita de que la aserción y el modo indicativo puedan identificarse tan próximamente entre sí. Y esto es porque hay muchas emisiones de oraciones indicativas que no son aserciones, por ejemplo las oraciones indicativas emitidas como juego, simulación, broma y ficción; y por supuesto que pueden hacerse aserciones emitiendo oraciones en otros modos. (Emisiones de "¿Notaste que Susana lleva nuevamente su sombrero violeta?" o "Fíjate que Susana lleva nuevamente su sombrero violeta", en ocasiones pueden ser simplemente aserciones de que Susana lleva nuevamente su sombrero violeta.) E igualmente para los demás modos; podemos hacer una pregunta con una indicativa o una imperativa ("Dime quien ganó la tercera carrera", "Quisiera saber tu número de teléfono"), o efectuar una orden con una indicativa ("En esta casa nos quitamos los zapatos antes de entrar").

Es innecesario decir que Dummett sabe todo esto, y si provisoriamente se permite pasar por alto estos casos en favor de una idea mayor, es sólo porque cree que en un sentido claro los contraejemplos producen desviaciones. ¿Pero cuál es este sentido? Austin hizo una distinción entre lo que él llamó los usos "normales" o "serios" de una oración y los usos "descoloridos" o "parásitos".[4] Si una distinción de este tipo pudiera hacerse de una manera no circular, y resultara así que el uso normal o serio de las indicativas fuera hacer aserciones, el de las imperativas dar órdenes, el de las interrogativas hacer preguntas y así sucesivamente, entonces se habría establecido la conexión deseada entre los modos y los usos de las oraciones.

[3] M. Dummett, *Frege: Philosophy of Language*, 315, 316.
[4] J. L. Austin, *How to Do Thing with Words*, 22.

Seguramente hay alguna conexión importante entre los modos y sus usos, y eso nos hace pensar que hay algo natural, serio o normal en el uso de una oración en un cierto modo para ejecutar un acto "correspondiente". La cuestión es si esta sensación puede articularse en una forma que arroje luz sobre la naturaleza de los modos. Es fácil ver que apelar a lo "serio" o "normal" no va más allá de una apelación a la intuición. El que una orden se emita en el imperativo en vez del indicativo no nos da ninguna clave de su seriedad; igualmente, una pregunta seria puede expresarse en el modo imperativo en vez del interrogativo. Y si "normal" significa usual, o estadísticamente más frecuente, resulta por cierto dudoso que la mayoría de las indicativas se emitan como aserciones. Hay demasiadas historias, repeticiones memorizadas, ilustraciones, suposiciones, parodias, charadas, salmos, y cumplidos conspicuamente impensados. Y en todo caso el análisis del modo no puede descansar de manera plausible sobre los resultados de este tipo de estudio estadístico.

La solución de Dummett consiste en pasar de lo serio o normal a lo convencional: una aserción es una indicativa emitida bajo condiciones especificadas por convención; una orden es una imperativa emitida bajo otras condiciones dadas convencionalmente, y así sucesivamente. Escribe entonces el autor: "... la aserción consiste en la emisión (deliberada) de una oración a la cual, por su forma y contexto, se reconoce como siendo usada de acuerdo con una cierta convención general..."[5] Y de las imperativas, "... la emisión de una oración de una forma determinada, a menos que circunstancias especiales alejen a este acto de su significado usual, constituye en sí misma dar una orden".[6] Agrega esta advertencia acerca de cómo aproximarse al tema de las relaciones entre los modos y sus usos:

> ...la aproximación correcta equivale a considerar que las emisiones están demarcadas convencionalmente en tipos según la forma de las expresiones lingüísticas empleadas, y luego a inquirir acerca de las convenciones que gobiernan el uso de los varios tipos de emisión.[7]

La opinión de Dummett de que las acciones lingüísticas como la aserción y la orden consisten en emitir oraciones en los modos indicativo o imperativo bajo condiciones especificadas convencionalmente es esencial para su idea del lenguaje junto con la tesis de que hay una convención adicional según la cual las aserciones se hacen con la intención de decir lo que es verdadero. Y esto porque la unión de estas dos ideas establecería una conexión directa entre los lenguajes tal como se los usa en formas convencionales y un cierto propósito global (decir lo que es verdadero).[8]

[5] M. Dummett, *Frege: Philosophy of Language*, 311.
[6] *Ibíd.*, 301, 302.
[7] *Ibíd.*, 302.
[8] Para una mayor discusión de las ideas de Dummett sobre la convención véase el Ensayo 8.

Concuerdo en que debemos encontrar conexiones entre la forma en que se usan las oraciones y lo que ellas significan si es que queremos dar una consideración fundacional del lenguaje. Tengo mis dudas, sin embargo, de que alguno de los eslabones de la cadena de Dummett sea válido. No puedo discutir ahora el segundo eslabón, la supuesta convención de que se trata de decir lo que es verdadero. Pero en el presente contexto es relevante tratar acerca de la presunción de que la emisión de una oración indicativa bajo condiciones convencionales constituye una aserción.

Hay una dificultad que es obvia pero puede ser superable: si es que vamos a lograr una explicación *general* de la aserción en estos términos, tendrá que haber convenciones que expliquen cómo se hacen aserciones emitiendo oraciones que no están en el modo indicativo. Pero quizá sea plausible que, si hay convenciones que relacionan indicativas y aserciones, haya convenciones adicionales que relacionen otros modos con las aserciones.

El problema es que las convenciones requeridas no existen. Desde luego, es verdad que si una indicativa se emite bajo las condiciones correctas se habrá hecho una aserción. Hasta podríamos llegar a especificar condiciones que son necesarias y suficientes para hacer una aserción: por ejemplo, yo pienso que para hacer una aserción un hablante debe representarse a sí mismo como creyendo lo que dice. Pero nada de esto sugiere que las condiciones sean de naturaleza convencional.

También se debe admitir que los intérpretes y hablantes de un lenguaje son generalmente capaces de determinar cuándo se ha hecho una aserción, y que esta capacidad es una parte esencial de su competencia lingüística. Más aun, el conocimiento de las convenciones lingüísticas y de otros tipos desempeña un papel clave en la producción y detección de aserciones. La moda, la actitud, el tono, la función, el rol y el gesto tienen, o pueden tener, aspectos convencionales, y todos estos elementos pueden hacer una contribución crucial a la fuerza de una emisión. Podemos admitir fácilmente todo esto sin aceptar que tan sólo siguiendo una convención las emisiones indicativas o imperativas se convierten en aserciones u órdenes.

Pienso que hay poderosas razones para rechazar la idea de que hacer una aserción (o dar una orden, o hacer una pregunta) es ejecutar una acción puramente convencional. Una de las razones es, como lo he estado sugiriendo, que resulte tan difícil decir cuál es la convención. (Por ejemplo, si un asertor necesariamente se representa a sí mismo creyendo lo que dice, uno tendría que describir las convenciones siguiendo la forma en que uno puede representarse a sí mismo creyendo lo que uno dice.) Una segunda cuestión es la siguiente. Muy a menudo comprendemos una emisión en todos sus aspectos relevantes, excepto en que no sabemos si es o no una aserción. Una manera de engañar consiste en dejar abierto en la mente del engañado el tema de la aserción; las novelas históricas, o los *romans-à-clef*, nos dejan deliberadamente confundidos. ¿Se omite algún aspecto convencional de la aserción? ¿Cuál? Y

si pudiéramos decirlo, entonces ¿por qué el engañador o el escritor de estas novelas no incluiría ese mismo ítem en su emisión?

Todo aquello acerca de la aserción que sea convencional puede convertirse en palabras, o convertirse de algún modo en una parte explícita de la oración. Supongamos ahora que este no es el caso, por lo cual el signo de aserción de Frege no es precisamente el equivalente formal del modo indicativo, sino una expresión más completa del elemento convencional de la aserción. Es fácil ver que el mero decir una oración en ese modo reforzado no implica que de ello resulte una aserción; todo bromista, embustero y actor aprovecharía de inmediato el modo reforzado para simular la aserción. No tiene sentido, entonces, hablar del modo reforzado; el indicativo que tenemos disponible hace todo lo que el lenguaje puede hacer al servicio de la aserción. Pero dado que el indicativo no es tan poderoso como para que su mero empleo constituya una aserción, lo que debe agregarse para producir aserción no puede ser una simple cuestión de convención lingüística.

Lo que este argumento ilustra es una cualidad básica del lenguaje, que puede llamarse autonomía del significado lingüístico. Una vez que un rasgo del lenguaje ha obtenido expresión convencional, puede ser usado para servir a muchos fines extralingüísticos; la representación simbólica necesariamente rompe toda relación próxima con propósitos extralingüísticos. Aplicado al presente caso, esto significa que no puede haber una forma de discurso que, sólo a fuerza de su significado convencional, pueda usarse solamente para un propósito dado, tal como hacer una aserción o formular una pregunta.

El argumento tiene una forma simple: el modo no es un signo convencional de la aserción o de la orden, pues nada es, o podría ser, un signo convencional de la aserción o de la orden. La razón de esto, debe recalcarse, no reside en que la fuerza ilocutoria de un acto de habla sea un aspecto puramente mental, interior o intencional del acto.[9] Por supuesto que la aserción o la orden debe ser intencional, como debe serlo el significado en el sentido restringido. Pero es parte de la intención el que el acto deba ser interpretado como asertivo o imperativo y, por lo tanto es también parte de la intención el que algo públicamente aparente deba invitar a la interpretación apropiada.

Sería fácil verse envuelto en una disputa acerca del grado en que la intención de un hablante de ejecutar un acto que será interpretado como asertivo debe verificarse antes de que su acto sea correctamente identificado como una aserción. Es excesivo insistir en que una aserción se ha hecho sólo si es realmente interpretada como una aserción; es demasiado poco requerir solamente que la intención esté presente. No necesitamos dejar establecida la cuestión de hasta qué punto un asertor debe tener éxito en su intención; todo lo que aquí interesa es si un asertor o quien da una orden debe pretender que su oyente reconozca su intención a través del empleo que da a lo que conoce o cree ser una con-

[9] Debo mi agradecimiento a Dummett por hacerme apreciar esta cuestión.

vención lingüística. Si hubiera tal convención, tendría que resultarnos fácil decir qué es y, en la gran mayoría de los casos, decir si se le ha observado o no. Pero si bien podemos determinar usualmente si se ha hecho o no una aserción, en general no podemos decir qué convención se ha seguido. La razón de que no podamos decirlo es, como ya lo he manifestado, que no existe tal convención.

Sería un error concluir que no hay una conexión convencional entre los modos y sus usos. Sin duda no existiría tal conexión si ciertos análisis de los modos fueran correctos. David Lewis, por ejemplo, ha sugerido con audacia que todas las oraciones no indicativas pueden ser "tratadas como paráfrasis de los realizativos correspondientes, con la misma estructura de base, significado, intensión y valor de verdad".[10] Luego "Hierve ese huevo" equivaldría en análisis a "Yo ordeno que tú hiervas ese huevo". Lewis piensa que las dos oraciones podrían tener un rango de usos distinto, pero dado que esta diferencia, en su teoría, no surgiría de una diferencia de significado, su teoría simplemente niega que el modo tenga alguna significatividad convencional.

Un análisis con igual consecuencia fue propuesto muchos años atrás por Herbert Bohnert.[11] Este proponía que los imperativos tienen la estructura de cierto tipo de disyunciones. Así "Hierve ese huevo" se transformaría en "O hierves ese huevo, o X sucederá", donde X es algo que se presume indeseado por la persona a la cual se habla.

Estas teorías encuentran su fuerza en el hecho de que podemos (y a menudo lo hacemos) usar indicativos para efectuar la tarea que según cree Dummett convencionalmente se asigna a los otros modos. Y en efecto si queremos ir en esa dirección, disponemos de una teoría más simple y a mi parecer mejor, cual es asignar a los imperativos el mismo análisis semántico que se asigna al indicativo más directamente correspondiente (trátese a "Hierve ese huevo" tal como se trata a "Tú hervirás ese huevo").[12]

Una virtud de estas teorías es que tornan evidente el hecho de que tener un valor de verdad no es obstáculo para que se use una oración para dar una orden o para formular una pregunta. Pero este mérito de las teorías reductivas también explica su fracaso, pues el simple reducir los imperativos o los interrogativos a indicativos nos deja sin considera-

[10] D. Lewis, "General Semantics", 208.

[11] H. Bohnert, "The Semiotic Status of Commands".

[12] Las interrogativas sí-no serán tratadas entonces, quizá, como poseedoras de la misma semántica que la afirmativa indicativa correspondiente; o, según otra opción, como poseedoras de la misma semántica que la alternación de la afirmativa indicativa con la negación de la afirmativa indicativa. Tendría que atribuirse a las preguntas *WH** la misma semántica que las correspondientes oraciones abiertas en el indicativo. Aquí, al igual que en todo este artículo, mis afirmaciones respecto del modo interrogativo sólo están bosquejadas. Tal como me lo señalara Jaakko Hintikka, mi programa general para los modos puede encontrarse en problemas cuando se haga un intento serio de aplicarlo a las interrogativas.

* Las preguntas *WH* a las que hace referencia el autor son aquellas que introducen la interrogación con la palabras *who, what, where,* etc. [T.]

ción alguna de las diferencias entre los modos. Si alguna de las teorías reductivas es correcta, entonces, el modo es tan irrelevante para el significado como a menudo se dice que lo es la voz. Si el modo no afecta el significado, ¿cómo podemos tener esperanzas de explicar la conexión entre modo y uso, sea cual fuere esa conexión? Los análisis reductivos, más que resolver el problema con el cual comenzamos, lo abandonan.

Estamos ahora en posición de enumerar las características que debe tener una teoría del modo satisfactoria.

(1) Debe mostrar o preservar las relaciones entre los indicativos y las oraciones correspondientes en los otros modos; debe, por ejemplo, articular el sentido en el cual "Te quitarás los zapatos", "Quítate los zapatos" y "¿Te quitarás los zapatos?" tienen un elemento común.

(2) Debe asignar un elemento de significado a las emisiones en un modo dado que no esté presente en emisiones en otros modos. Y este elemento debe conectarse con la diferencia de fuerza entre aserciones, preguntas y órdenes de tal manera que explique nuestra intuición de una relación convencional entre modo y uso.

(3) Finalmente, la teoría debe ser semánticamente tratable. Si la teoría responde a los requerimientos de una teoría de la verdad, entonces diré que todo está bien. Y por el otro lado si, como creo que sostuvo Bar-Hillel, puede demostrarse que una teoría estándar de la verdad es incapaz de explicar el modo, entonces la teoría de la verdad es inadecuada como teoría general del lenguaje.

La dificultad de alcanzar los tres requisitos es obvia. Las primeras dos condiciones sugieren que el modo debe estar representado por operadores que gobiernan oraciones, siendo las oraciones gobernadas indicativas (en cuyo caso no se necesita un operador para el modo indicativo) o neutrales (en cuyo caso se necesita un operador para cada modo). La tercera condición, sin embargo, parece prohibir todos los operadores menos los oracionales veritativos funcionales, y está claro que los operadores veritativos funcionales no pueden servir para dar una interpretación plausible del modo.

Me parece que Dummett está en lo cierto cuando dice que la expresión sintáctica del modo no es igual a una expresión funcional, que no podemos inquirir cuál es su sentido de referencia, y que (entonces) una oración con un indicador de modo "ni expresa nada ni sostiene nada". Como dice Geach, un indicador de modo no es igual a ninguna otra parte del habla; "...es necesariamente *sui generis*. Pues cualquier otro signo lógico, si no es superfluo, modifica de alguna manera el contenido de una proposición; en tanto que éste no modifica el contenido...".[13]

Dummett y Geach hacen estas declaraciones negativas a partir de lo que a mí me parecen razones parcialmente equivocadas o confusas. Dummett piensa que una oración con un indicador de modo no puede expresar o sostener nada porque la oración "...afirma algo... esto es, que

[13] P. T. Geach, "Assertion", 458.

el pensamiento expresado por lo que sigue al rasgo de juicio es verdadero".[14] Yo he insistido en que esto es algo que ninguna expresión puede hacer; pero la idea también parece equivocada por otra razón. Si el signo de aserción afirma que el pensamiento expresado por el resto de la oración es verdadero, luego el signo imperativo debe afirmar que el pensamiento expresado por el resto de la oración es verdadero, pasará a ser verdadero. Pero esta proposición elimina la distinción entre aserción y orden. Geach dice en cambio que el indicador de modo (entendido tal como lo entendió Frege) "...muestra que la proposición está siendo afirmada".[15] Esta proposición preserva la distinción necesaria.

He argumentado en contra tanto de Geach como de Dummett, que ningún indicador de modo puede mostrar o afirmar o de ningún otro modo determinar convencionalmente qué fuerza tiene su emisión. Pero si esto es así, no nos quedamos con ninguna idea clara acerca de cuál es el modo que contribuye al significado. Ciertamente parece que nos encontramos ante una paradoja. El modo debe contribuir de alguna manera al significado (punto 2 más atrás), pues resulta claro que el modo es una característica convencional de las oraciones. Sin embargo, no puede combinarse con o modificar el significado del resto de la oración en ninguna forma conocida.

Vayamos en busca del auxilio de los que Austin dio en llamar "realizativos explícitos". Hemos rechazado la idea presentada por David Lewis de que los imperativos sean reducidos a realizativos explícitos, pero ella permanece abierta para explotar analogías. Austin llamó la atención sobre el hecho de que "...en ocasiones podemos usar la emisión 'Ve' para obtener prácticamente lo mismo que obtenemos con la emisión 'Te ordeno ir' ".[16] Pero ¿cómo serán analizados los realizativos explícitos?

Austin sostuvo que los realizativos no tienen valor de verdad en base a que emitir una oración como "Te ordeno ir" no equivale típicamente a describir el acto de habla propio de uno sino más bien a dar una orden. Esta es tal vez una explicación exacta del modo en que caracterizaríamos muchos actos de habla que consisten en emitir realizativos explícitos. Pero como explicación de lo que significan las palabras que se emiten, este punto de vista introduce una discrepancia intolerable entre la semántica de ciertos verbos en primera persona del tiempo presente y sus variantes en otras personas de otros tiempos. Y el problema es adventicio, puesto que lo que resulta especial en los realizativos explícitos se explica mejor como consecuencia de un uso especial de palabras con un significado común que como consecuencia de un significado especial.

Si aceptamos cualquiera de las semánticas usuales para los realizativos explícitos, sin embargo, la dificultad recurre en una forma que es difícil de evitar. Según las consideraciones usuales del asunto, en una

[14] M. Dummett, *Frege: Philosophy of Language*, 316.
[15] P.T. Geach, "Assertion", 458.
[16] J. L. Austin, *How to Do Thing with Words*, 32.

oración como "Jones ordenó a Smith que fuera" las palabras finales ("a Smith que fuera") sirven para nombrar o describir una oración, o una proposición, o el sentido de una oración. Para mostrar la oración anidada relevante, podemos remodelar el todo así: "Jones ordenó a Smith que fuera el caso que Smith va". Y ahora, para las consideraciones usuales, la oración "Smith va" no puede, en este contexto, tener nada que ver con su significado ordinario. Por lo tanto, tampoco puede tener nada que ver con su espectro de usos ordinario. Pero "Te ordeno ir" (o, reconstruida, "Te ordeno que sea el caso que tú vas") tiene la misma forma que "Jones ordenó a Smith que fuera", y entonces debería tener el mismo análisis con los cambios apropiados de persona y tiempo. Se sigue que al emitir "Te ordeno ir" yo no puedo significar con las palabras "tú vas" nada parecido a lo que significaría con ellas si estuvieran solas; en el contexto presente, estoy usando estas palabras meramente para referir una oración o la proposición que ella expresa. Parecería imposible, entonces, que si cualquier análisis usual de dichas palabras es correcto, una emisión de "Te ordeno ir" pudiera ser una orden de ir.

O, para decir lo mismo respecto de la aserción: una forma de establecer el hecho de que no estoy afirmando que está lloviendo cuando emito las palabras "Está lloviendo" es prefijar las palabras "Jones afirmó *that*". De acuerdo con la mayoría de los análisis de tales oraciones, debe esperarse el mismo efecto si prefijo las palabras "Yo afirmo *that*".

Esta dificultad es una entre las dificultades de los análisis usuales que me han llevado a urgir una aproximación enteramente distinta a la semántica del discurso indirecto, de las oraciones de creencia, de las oraciones acerca de órdenes, esperanzas, expectativas, y así sucesivamente: todo el profano abanico de las locuciones atributivas de actitudes.[17] Dejando de lado las complicaciones que surgen cuando cuantificamos desde afuera las oraciones gobernadas (o sus confusas superficies), mi propuesta es la siguiente. Al tener en mente que en todo caso son emisiones, no oraciones, las que tienen un valor de verdad y una semántica específicos, deberíamos estar satisfechos con un análisis de las condiciones de verdad de emisiones de palabras como "Jones afirmó *that* está lloviendo". Sugiero que consideremos tal emisión como la emisión de dos oraciones: "Jones afirmó *that*" y luego "Está lloviendo". Si yo afirmo que Jones afirmó que está lloviendo, lo hago cuando afirmo "Jones afirmó *that*" y luego emito, usualmente en forma no asertiva, la oración que da el contenido de la aserción de Jones; en este caso, "Está lloviendo". La función del "*that*" en una emisión de "Jones afirmó *that*" es referir a la emisión siguiente, la cual da el contenido. Así, para expresar la idea en una forma recargada pero sugestiva: una emisión de "Jones afirmó *that* está lloviendo" tiene el efecto de dos emisiones:

Jones hizo una aserción cuyo contenido está dado por mi próxima emisión. Está lloviendo.

[17] Véase el Ensayo 7.

Este análisis da cuenta del habitual fracaso de la sustitutividad en las atribuciones de actitud que no invocan alguna semántica no estándar, pues la referencia del "*that*" cambia con cualquier cambio en la emisión siguiente. También permite que la segunda emisión consista, en ocasiones, en hacer una aserción, como lo hará si yo digo realmente, "Hago una aserción cuyo contenido está dado por mi próxima emisión". De manera similar, puedo estar dando una orden al decir "Vas tú" aun si estas palabras siguen a una emisión de "Yo ordeno *that*", o "Esto es una orden".

Propongo tratar a los modos no indicativos de una manera muy parecida a como se trata a los realizativos explícitos, pero no reduciendo los otros modos al indicativo. He aquí la idea. Bien podemos dejar solos a los indicativos, puesto que no hemos encontrado ningún uso inteligible para un signo de aserción. Seguiremos adelante, como es nuestra costumbre, a veces usando las oraciones indicativas para hacer aserciones, a veces usándolas para hacer otras cosas; y continuaremos usando oraciones en otros modos para hacer aserciones cuando podamos y lo encontremos atractivo.

En inglés se señalan los modos no indicativos en varias formas, ocasionalmente ambiguas, mediante cambios en el verbo, en el orden de las palabras, en la puntuación o en la entonación. Podemos pensar, entonces, que las oraciones no indicativas son oraciones indicativas más una expresión que representa sintácticamente la transformación apropiada; llamemos a esta expresión el *fijador de modo*. Y así como una oración no indicativa puede descomponerse en una oración indicativa y un fijador de modo, del mismo modo una emisión de una oración no indicativa puede descomponerse en dos actos de habla distintos, uno la emisión de una oración indicativa, y el otro la emisión de un fijador de modo. No debemos preocuparnos porque de hecho usualmente no ejecutemos estos actos uno tras otro sino más o menos simultáneamente. Sólo piénsese en alguien que se restriega el vientre con una mano y se rasca la cabeza con la otra.

Hemos visto que el fijador de modo no puede ser tratado semánticamente como un operador de modo de ningún tipo común, y que parecería imposible ofrecer una consideración admisible de cómo el significado de una oración no indicativa puede ser el resultado de combinar el significado de una indicativa con el significado del fijador de modo. Sugiero que aceptemos la independencia semántica entre las indicativas y los fijadores de modo que las acompañan evitando incorporar el fijador de modo a una oración simple en el indicativo. Está la oración indicativa por un lado, y luego, o al mismo tiempo, el fijador de modo. O todavía mejor, si pensamos en la emisión, está la emisión del elemento indicativo, y está (quizá simultáneamente) la emisión del fijador de modo. La emisión de una no indicativa puede de esta manera descomponerse siempre en la ejecución de dos actos de habla.

Hasta ahora, la propuesta no es claramente compatible con las propuestas de Geach, Dummett, y tal vez otros. Por cierto, he dejado de lado al signo de aserción, pero eso puede considerarse mayormente una

cuestión notacional. También he descartado una explicación del significado del operador de modo en términos de un indicador convencional de la fuerza con la cual se hace una emisión particular. Se produce así un vacío en el centro de mi descripción; no he podido decir qué significa el indicador de modo.

Geach señaló que lo que yo llamo fijador de modo no puede verse como cualquier otra parte del habla. Esto obedece a que él lo interpretó como una parte de una oración más larga, y se dio cuenta de que no tenía las propiedades semánticas de un operador oracional. Hemos extraído al fijador de modo de la oración indicativa que acompaña. La única forma que el fijador de modo puede tener —la única función que puede ejecutar— es la de una oración. Se comporta como una oración, de la cual una emisión refiere una emisión de una oración indicativa. Si fuéramos a representar en forma lineal la emisión de, digamos, la oración imperativa "Ponte el sombrero", resultaría ser la emisión de una oración como "Mi próxima emisión es imperativa en cuanto a su fuerza", seguida por una emisión de "Te pondrás el sombrero".

Esto da una idea de la situación semántica, pero la sintaxis no funciona. El fijador de modo no puede ser ninguna oración real del castellano, puesto que representa una cierta transformación. Yo no quiero afirmar que las oraciones imperativas sean dos oraciones indicativas. Más bien, podemos dar la semántica de la emisión de una oración imperativa si consideramos los especificaciones de condiciones de verdad, las condiciones de verdad de la emisión de una oración indicativa obtenida al transformar la imperativa original, y las condiciones de verdad del fijador de modo. El fijador de modo de una emisión de "Ponte el sombrero" es verdadero si y sólo si la emisión del núcleo indicativo es imperativa en cuanto a su fuerza.

Los fijadores de modo caracterizan a una emisión en cuanto poseedora de una cierta fuerza ilocutoria; no afirman que ella tenga esa fuerza, puesto que solamente los hablantes hacen aserciones. Pero si alguien desea dar una orden, bien puede hacerlo emitiendo el fijador de modo imperativo asertivamente. Luego si la condiciones de verdad del fijador de modo son válidas (si lo que el hablante ha afirmado es verdadero), su emisión del núcleo indicativo constituirá el dar una orden. Hay gran cantidad de otras formas en que ese alguien puede dar la misma orden; por ejemplo, si afirma "Esto es una orden", o "En este acto ordeno que"; o simplemente si emite "Te quitarás el sombrero" como una orden. Creo que esta propuesta satisface los tres requisitos que enumeramos para un análisis satisfactorio de los modos.

Primero, en la propuesta hay un elemento común a los modos. Sintácticamente consiste en el núcleo indicativo, que está transformado en los modos no indicativos. Semánticamente consiste en las condiciones de verdad de este núcleo indicativo.

Segundo, el modo está representado sistemáticamente por el fijador de modo (o por su ausencia en el caso de indicativo). Los fijadores de modo funcionan semánticamente como oraciones, de las cuales hay emisiones que son verdaderas o falsas según la emisión del núcleo indicati-

vo, tenga o no tenga la fuerza ilocutoria especificada. El significado del fijador de modo es convencional en cualquier sentido en que el significado en general lo sea, pero no se sugiere que este significado determine la fuerza ilocutoria de una emisión del fijador de modo, de su indicativa asociada, o del par. La conexión convencional entre modo y fuerza es más bien la siguiente: el concepto de fuerza es parte del significado de modo. Una emisión de una oración imperativa, en efecto, dice de sí misma que tiene una cierta fuerza. Pero este no es el "dice" del "afirma" (excepto en ocasiones y como adición). Lo que ella dice, en este sentido no afirmado, puede ser tanto falso como verdadero. Este hecho no afecta la conexión conceptual entre modo y fuerza.

Tercero, una semántica frontal, basada en una teoría de la verdad para las emisiones, funcionará tan bien aquí como en cualquier otra parte. En particular, todas las emisiones que la teoría considera básicas tienen un valor de verdad en el sentido habitual. Por otra parte, si yo estoy en lo cierto, no puede decirse que una emisión de una oración no indicativa tenga un valor de verdad. Pues cada emisión de una no indicativa tiene su fijador de modo, y por lo tanto desde el punto de vista semántico debe considerarse que consiste en dos emisiones. Cada una de las dos emisiones tiene un valor de verdad, pero la emisión combinada no es la emisión de una conjunción, y entonces no tiene un valor de verdad.

Interpretación radical

9

Interpretación radical

Kurt emite las palabras *"Es regnet"* y bajo las condiciones correctas nosotros sabemos que ha dicho que está lloviendo. Si identificáramos su emisión como intencional y lingüística, somos capaces de seguir adelante e interpretar sus palabras: podemos decir lo que sus palabras, en esa ocasión, significaban. ¿Qué podríamos saber que nos ayudaría a lograr esto? ¿Cómo podríamos haberlo sabido? La primera de estas preguntas no es igual a la pregunta acerca de que *sí* sabemos que nos permite interpretar las palabras de otros. Pues fácilmente puede haber algo que podríamos saber y no sabemos, cuyo conocimiento bastaría para la interpretación, mientras que por otra parte no es completamente obvio que haya algo que realmente sepamos y que desempeñe un papel esencial en la interpretación. La segunda pregunta, cómo podríamos haber llegado a tener un conocimiento que nos sirviera para obtener interpretaciones, naturalmente no tiene relación con la historia real de la adquisición del lenguaje. Es así una pregunta doblemente hipotética: dada una teoría que haría posible la interpretación, ¿qué evidencia plausiblemente disponible para un intérprete potencial apoyaría la teoría hasta un punto razonable? A continuación trataré de pulir estas preguntas y sugerir respuestas.

El problema de la interpretación es tan foráneo como doméstico: surge entre los hablantes de un mismo lenguaje bajo la forma de la pregunta ¿cómo puede determinarse que el lenguaje es el mismo? Los hablantes del mismo lenguaje pueden suponer que para ellos las mismas expresiones deben ser interpretadas de la misma manera, pero esto no indica que justifica la suposición. Toda comprensión del discurso de otro involucra la interpretación radical. Pero para ayudar a evitar que las suposiciones pasen desapercibidas centraremos nuestra atención en los casos donde más claramente se requiere interpretación: la interpretación en un idioma del habla en otro idioma.[1]

¿Qué conocimiento serviría para la interpretación? Una respuesta breve sería: conocimiento de lo que significa cada expresión significati-

[1] El término "interpretación radical" está pensado para sugerir un fuerte parentesco con la "traducción radical" de Quine. Parentesco no es identidad, sin embargo, e "interpretación" en lugar de "traducción" señala una de las diferencias: un mayor énfasis en lo explícitamente semántico en el primero.

va. En alemán, esas palabras que dijo Kurt significan que está lloviendo y Kurt estaba hablando alemán. Así, al emitir las palabras *"Es regnet"* Kurt dijo que estaba lloviendo. Esta respuesta no es una mera reformulación del problema, como en principio podría pensarse. No lo es porque sugiere que al pasar de una descripción que no interpreta (su emisión de las palabras *"Es regnet"*) a la interpretación de esa descripción (su decir que está lloviendo) debemos introducir una maquinaria de palabras y expresiones (que puede estar o no ejemplificada en emisiones reales), y esta sugerencia es importante. Pero la respuesta no da más ayuda que ésa, pues no dice qué es saber lo que significa una expresión.

Por cierto, también existe la insinuación de que en correspondencia con cada expresión significativa que sea una entidad está su significado. Esta idea, si bien no es errónea, ha probado ser de muy poca utilidad: a lo sumo hipostatiza el problema.

La propagación de cierto desencanto hacia los significados en cuanto implementadores de una explicación viable de la comunicación o de la interpretación ayuda a explicar por qué algunos filósofos han tratado de arreglárselas dejando de lado no sólo los significados, sino también todo esbozo de una teoría seria. Cuando los conceptos que convocamos para tratar de explicar la interpretación demuestran ser más desconcertantes que el *explanandum* resulta tentador reflexionar que después de todo la comunicación verbal consiste nada más que en elaboradas perturbaciones en el aire que forman una ligazón causal entre las actividades no lingüísticas de los agentes humanos. Pero si bien los dichos interpretables no son sino (es decir son idénticos a) acciones ejecutadas con variadas intenciones no lingüísticas (advertir, controlar, entrenar, distraer, insultar), y a su vez estas acciones no son sino (son idénticas a) movimientos intencionales de los labios y de la laringe, esta observación no nos aproxima en lo más mínimo a una consideración general inteligible de lo que tendríamos que saber para poder redescribir las emisiones ininterpretadas como emisiones interpretadas correctas.

El recurso a los significados nos deja más desamparados que cuando comenzamos a partir de los hechos no lingüísticos que deben proveer la base probatoria para la interpretación; la postura del "no son sino" no proporciona ninguna clave de la forma en que la evidencia se relaciona con aquello para lo cual seguramente ella es evidente.

Otras propuestas para salvar el escollo se quedan por el camino en distintas formas. Las teorías "causales" de Ogden y Richards y de Charles Morris trataron de analizar el significado de las oraciones, tomando una por vez, en base a los datos conductistas. Aun si estas teorías hubieran funcionado para las oraciones más simples (cosa que indudablemente no hicieron), no llegaron a tocar el problema de extender el método a oraciones de mayor complejidad y abstracción. Otros tipos de teorías parten de intentar conectar palabras, y no oraciones, con hechos no lingüísticos. Esto resulta prometedor pues el número de palabras es finito en tanto que el de oraciones no lo es, y sin embargo cada oración no es más que una concatenación de palabras: esto ofrece la posibilidad de una teoría que interprete cada una de las infinitas oraciones usando

solamente recursos finitos. Pero estas teorías no logran llegar a la evidencia, pues parece claro que las características semánticas de las palabras no pueden explicarse directamente sobre la base de los fenómenos no lingüísticos. La razón es simple. Los fenómenos hacia los que debemos dirigirnos son los intereses y actividades extralingüísticos a los que sirve el lenguaje, y éstos son satisfechos por palabras sólo en tanto y en cuanto las palabras estén incorporadas en (o en ocasiones resulten ser) oraciones. Pero entonces no hay posibilidad de dar una apreciación fundacional de las palabras sin dar antes una de las oraciones.

Por razones muy diferentes, la interpretación radical no puede tener esperanzas de tomar como evidencia del significado de una oración a una explicación de las complejas y delicadamente discriminadas intenciones con las cuales se emite típicamente la oración. No es fácil ver cómo una aproximación así puede habérselas con el rasgo estructural y recursivo del lenguaje que resulta esencial para explicar la forma en que serán entendidas las nuevas oraciones. Pero la dificultad central es que no podemos esperar adjudicar un sentido a la atribución de intenciones finamente discriminadas independientemente de la interpretación del discurso. La razón no es que no podamos formular preguntas necesarias, sino que interpretar las intenciones de un agente, sus creencias y sus palabras son partes de un único proyecto, ninguna de cuyas partes puede suponerse completa antes de que el resto lo esté. Si esto es correcto, no podemos convertir a toda la panoplia de intenciones y creencias en la base probatoria para una teoría de la interpretación radical.

Estamos ahora en posición de decir algo más acerca de lo que serviría para hacer posible la interpretación. El intérprete debe ser capaz de comprender cualquiera de las infinitas oraciones que el hablante podría emitir. Si queremos establecer explícitamente lo que el intérprete debería saber para estar en condiciones de hacerlo, debemos expresarlo en forma finita.[2] Para que este requerimiento pueda cumplirse debe abandonarse toda esperanza de un método universal de interpretación. Lo máximo que puede esperarse es explicar cómo un intérprete podría interpretar las emisiones de hablantes de un único lenguaje (o de un número finito de lenguajes): no tiene sentido pedir una teoría que diera por fruto una interpretación explícita para cualquier emisión en cualquier (posible) lenguaje.

Todavía no está claro, desde luego, qué es para una teoría dar una interpretación explícita de una emisión. La formulación del problema parece invitarnos a pensar que la teoría es la especificación de una función que toma a las emisiones como argumentos y que tiene a las interpretaciones como valores. Pero entonces las interpretaciones no serían mejores que los significados y casi con seguridad no serían mejores que entidades de algún tipo misterioso. Luego parece sensato describir lo que se quiere de la teoría sin hacer referencia aparente a significados o

[2] Véase el Ensayo 1.

interpretaciones: alguien que conozca la teoría puede interpretar las emisiones para las cuales la teoría es aplicable.

El segundo requerimiento general de una teoría de la interpretación es que se la puede apoyar o verificar mediante evidencia plausiblemente disponible para un intérprete. Dado que la teoría es general —debe aplicarse a un potencial infinito de emisiones— sería natural pensar que la evidencia en su favor consista en instancias de interpretaciones particulares reconocidas como correctas. Y este caso, por supuesto, se presenta al intérprete que trata con un lenguaje que ya conoce. El hablante de un lenguaje normalmente no puede producir una teoría finita explícita para su propio lenguaje, pero puede poner a prueba una teoría propuesta dado que puede determinar si ella ofrece interpretaciones correctas cuando se la aplica a emisiones particulares.

En la interpretación radical, sin embargo, se supone que la teoría provee una comprensión de emisiones particulares que no está dada de antemano, por lo que la evidencia última para la teoría no puede consistir en interpretaciones correctas de muestras. Para entenderse con el caso general, la evidencia debe ser de una clase que estaría a disposición de alguien que aún no sabe cómo interpretar las emisiones que la teoría está preparada para cubrir: debe tratarse de evidencia que pueda establecerse sin un uso fundamental de conceptos lingüísticos tales como significado, interpretación, sinonimia y similares.

Antes de decir qué tipo de teoría me parece que funcionaría bien, quiero discutir una última sugerencia alternativa, cual es que toda la teoría que necesitamos se limitaría a un método de traducción del lenguaje a ser interpretado al lenguaje del intérprete. Una teoría así consistiría en el enunciado de un método efectivo para pasar de una oración arbitraria de la lengua extranjera a una oración de un lenguaje familiar; de este modo satisfaría el requerimiento de un método finitamente enunciado aplicable a cualquier oración. Pero a mí no me parece que la mejor forma que pueda tomar una teoría de interpretación sea la de un manual de traducción.[3]

Cuando nuestra meta es la interpretación, un método de traducción se ocupa de un tópico equivocado, una relación entre dos lenguajes, mientras que lo que se busca es una interpretación de uno (en otro, por supuesto, pero eso no hace falta decirlo ya que toda teoría está en algún lenguaje). No podemos, sin confundirnos, contar al lenguaje empleado en la enunciación de la teoría como parte de la materia de que trata la teoría a menos que explícitamente lo hagamos así. En el caso general, una teoría de traducción involucra tres lenguajes: el lenguaje objeto, el

[3] La idea de un manual de traducción con restricciones empíricas apropiadas como recurso para estudiar problemas en la filosofía del lenguaje es, por supuesto, de Quine. Esta idea inspiró buena parte de mi pensamiento sobre el tema presente, y mi propuesta es muy cercana a la de Quine en aspectos importantes. Puesto que Quine no pretendió responder a las preguntas que he establecido, la afirmación de que el método de traducción no es adecuado como solución al problema de la interpretación radical no representa una crítica a ninguna doctrina de Quine.

lenguaje sujeto y el metalenguaje (los lenguajes del cual y hacia el cual pasa la traducción, y el lenguaje de la teoría, que dice cuáles expresiones del lenguaje sujeto traducen cuáles expresiones del lenguaje objeto). Y en este caso general, podemos saber cuáles oraciones del lenguaje sujeto traducen cuáles oraciones del lenguaje objeto sin conocer qué significa cualquiera de las oraciones de ambos lenguajes (en algún sentido, de todas maneras, eso permitiría a alguien que comprendiera la teoría interpretar oraciones del lenguaje objeto). Si el lenguaje sujeto resulta ser idéntico al lenguaje de la teoría, luego alguien que comprenda la teoría puede sin duda usar el manual de traducción para interpretar emisiones extrañas; pero esto tiene lugar porque esa persona se vale de dos cosas que conoce y que la teoría no enuncia: el hecho de que el lenguaje sujeto es también el suyo, y su conocimiento para interpretar emisiones en su propio lenguaje.

Es muy dificultoso tratar de hacer explícita la suposición de que una oración mencionada pertenece al lenguaje propio de uno. Podríamos intentarlo, por ejemplo, con " *'Es regnet'* en el lenguaje de Kurt se traduce como 'Está lloviendo' en el mío", pero la autorreferencia deíctica está fuera de lugar en una teoría que se propone funcionar para cualquier intérprete. Si decidimos aceptar esta dificultad, todavía resta el hecho de que el método de traducción deja que permanezca tácitamente y fuera del alcance de la teoría aquello que necesitamos conocer para poder interpretar nuestro propio lenguaje. Una teoría de la traducción debe descubrir algún tipo de estructura dentro de las oraciones, pero no hay razón alguna para esperar que nos proporcione una idea de la forma en que los significados de las oraciones dependen de sus estructuras.

Una teoría satisfactoria para interpretar las emisiones de un lenguaje, incluido el nuestro, revelará una estructura semántica significativa: por ejemplo, que la interpretación de emisiones de oraciones complejas depende de la interpretación de oraciones más simples. Supóngase que a una teoría de la traducción fuéramos a agregarle una teoría satisfactoria de la interpretación para nuestro propio lenguaje. Así tendríamos exactamente lo que queremos, pero en una forma innecesariamente voluminosa. El manual de traducción produce mecánicamente, para cada oración del lenguaje a traducir, una oración en el lenguaje del traductor; la teoría de la interpretación da luego la interpretación de estas oraciones familiares. Está claro que la referencia al idioma materno es superflua; es un intermediario no requerido entre la interpretación y el idioma extraño. Las únicas expresiones que tiene que mencionar una teoría de la interpretación son aquellas que pertenecen al lenguaje a ser interpretado.

Puede entonces verse una teoría de la interpretación para un lenguaje objeto como el resultado de la fusión de una teoría de la interpretación para un lenguaje conocido que sea estructuralmente reveladora, y de un sistema de traducción del lenguaje desconocido al conocido. La fusión hace que toda referencia al lenguaje conocido se vuelva ociosa; cuando esta referencia se deja de lado, lo que queda es una teoría de la interpretación estructuralmente reveladora para el lenguaje objeto

(apoyada, por supuesto, en palabras familiares). Contamos con estas teorías, a mi modo de ver, entre las teorías de la verdad del tipo que Tarski mostró por primera vez cómo producir.[4]

Lo que caracteriza a una teoría de la verdad del estilo de Tarski es que implica, para cada oración *s* del lenguaje objeto, una oración de la forma:

s es verdadera (en el lenguaje objeto) si y sólo si *p*.

Las instancias de la forma (que llamaremos oraciones T) se obtienen al reemplazar "*s*" por una descripción canónica de *s*, y "*p*" por una traducción de *s*. La noción semántica indefinida de importancia en la teoría es la de *satisfacción*, la cual relaciona oraciones, abiertas o cerradas, con secuencias infinitas de objetos, los cuales pueden considerarse como pertenecientes al rango de variables del lenguaje objeto. Los axiomas, que son finitos en número, son de dos tipos: algunos dan las condiciones bajo las cuales una secuencia satisface una oración compleja en base a las condiciones de satisfacción de oraciones más simples, otros dan las condiciones bajo las cuales se ven satisfechas las oraciones (abiertas) más simples. La verdad se define para las oraciones cerradas en términos de la noción de satisfacción. Una teoría recursiva como ésta puede convertirse en una definición explícita según lineamientos familiares, tales como lo muestra Tarski, siempre que el lenguaje de la teoría contenga suficiente teoría de conjuntos; pero no debemos preocuparnos por este paso adicional.

Si los nombres propios y las expresiones funcionales son características irreductibles del lenguaje objeto surgirán nuevas complicaciones. La cuestión se vuelve más engañosa en lo que respecta a los recursos deícticos. Tarski estaba interesado en los lenguajes formalizados que no contienen aspectos deícticos o demostrativos. Es así como pudo tratar a las oraciones como vehículos de la verdad; la extensión de la teoría a las emisiones es en este caso trivial. Pero los lenguajes naturales están indispensablemente repletos de rasgos deícticos, como los tiempos verbales, y entonces sus oraciones pueden variar su verdad según el tiempo y el hablante. El remedio es caracterizar la verdad para un lenguaje en relación a un tiempo y a un hablante. La extensión a las emisiones se hace nuevamente directa.[5]

Lo que sigue es una defensa del argumento de que una teoría de la verdad, modificada para su aplicación a un lenguaje natural, puede usarse como teoría de la interpretación. La defensa consistirá en intentos de responder a tres preguntas:

[4] A. Tarski, "The Concept of Truth in Formalized Languages".

[5] Para una discusión acerca de cómo puede una teoría de la verdad manipular demostrativos y cómo debe modificarse la Convención T, véase S. Weinstein, "Truth and Demonstratives".

1. ¿Es razonable pensar que una teoría de la verdad del tipo descrito puede formularse para un lenguaje natural?

2. ¿Sería posible decir que tal teoría es correcta en base a la evidencia plausiblemente disponible para un intérprete sin conocimiento previo del lenguaje a ser interpretado?

3. Si se supiera que la teoría es verdadera, ¿sería posible interpretar emisiones de hablantes del lenguaje?

La primera pregunta está dirigida a la suposición de que puede darse una teoría de la verdad para un lenguaje natural; la segunda y tercera preguntas inquieren acerca de si una teoría así satisfaría las demás exigencias que hemos requerido de una teoría de la interpretación.

1. *¿Puede formularse una teoría de la verdad para un lenguaje natural?*

Nos resultará útil para apreciar el problema considerar brevemente el caso en que un fragmento significativo de un lenguaje (más uno o dos predicados semánticos) se utiliza para enunciar su propia teoría de la verdad. De acuerdo con la Convención T de Tarski, un test de la adecuación de una teoría consiste en que ella implique todas las oraciones T. En apariencia, este test no puede cumplirse sin asignar a las oraciones del lenguaje algo muy parecido a una forma cuantificacional estándar, y sin apelar en la teoría a una noción relacional de satisfacción.[6] Pero lo que más impacta de las oraciones T es que sea cual fuere la maquinaria que debe funcionar para producirlas, y sean cuales fueren los engranajes ontológicos que deben girar, al final una oración T enuncia las condiciones de verdad de una oración usando recursos que no son más ricos que, pues son iguales a, aquellos de la oración misma. A menos que la oración original mencione mundos posibles, entidades intensionales, propiedades o proposiciones, el enunciado de sus condiciones de verdad no lo hace.

No hay una manera igualmente simple de expresar una idea análoga respecto de un lenguaje extraño sin apelar, como lo hace Tarski, a una noción inanalizada de traducción. Pero debemos ser capaces de hacer para otro lenguaje lo que hacemos para el nuestro; el problema con que nos toparemos será poder saber si lo estamos haciendo.

La restricción que impone la exigencia de una teoría que satisfaga la Convención T parece ser considerable: no se conoce hoy un método aceptado en general para tratar, dentro de esa restricción, con un espectro de problemas, por ejemplo, oraciones que atribuyen actitudes, modalidades, enunciados causales generales, contrafácticos, adjetivos atributivos, cuantificadores como "La mayoría" y otros por el estilo. Por otra parte, se ve lo que para mí es un progreso bastante notable. Para mencionar algunos ejemplos, allí están los trabajos de Tyler Burge sobre los

[6] Véase J. Wallace, "On the Frame of Reference", y el Ensayo 3.

nombres propios,[7] de Gilbert Harman sobre *"ought"*,[8] de John Wallace sobre los términos masa y los comparativos,[9] y mi propio trabajo sobre las atribuciones de actitudes y realizativos,[10] sobre los adverbios, sucesos, y enunciados causales singulares,[11] y sobre la cita.[12]

Si nos inclinamos por ser pesimistas respecto de lo que queda por hacer (¡o por parte de lo que ya se ha hecho!), deberíamos pensar en el magnífico logro de Frege al poner bajo control eso que Dummett llama "generalidad múltiple".[13] Frege no tenía en mente una teoría de la verdad en el sentido de Tarski, pero es obvio que persiguió, y encontró, estructurar de una clase para la cual puede establecerse una teoría de la verdad.

La tarea de aplicar en detalle una teoría de la verdad a un lenguaje natural en la práctica se dividirá casi con certeza en dos etapas. En la primera etapa se caracterizará la verdad, no para todo el lenguaje, sino para una parte del lenguaje cuidadosamente seleccionada. Esta parte, si bien será sin duda gramáticalmente incómoda, contendrá una infinidad de oraciones que agotarán el poder expresivo del lenguaje todo. La segunda parte apareará cada una de las oraciones restantes a una o (en caso de ambigüedad) más de una de las oraciones para las cuales se ha caracterizado la verdad. Podemos considerar que las oraciones para las que es aplicable la primera etapa de la teoría dan la forma lógica, o estructura profunda, de todas las oraciones.

2. *¿Puede verificarse una teoría de la verdad mediante el recurso a la evidencia disponible antes del comienzo de la interpretación?*

La Convención T dice que una teoría de la verdad es satisfactoria si genera una oración T para cada oración del lenguaje objeto. Basta con demostrar que una teoría de la verdad es empíricamente correcta, entonces, para verificar que las oraciones T son verdaderas (en la práctica, una muestra adecuada confirmará la teoría en un grado razonable). Las oraciones T sólo mencionan las oraciones cerradas del lenguaje, por lo que la evidencia relevante puede consistir exclusivamente en hechos acerca de la conducta y actitudes de los hablantes en relación con la oraciones (sin duda por medio de emisiones). Desde luego, una teoría que funcione debe tratar a las oraciones como concatenaciones de expresiones de longitud menor que la oracional, debe introducir nociones semánticas como satisfacción y referencia, y debe apelar a una ontología de secuencias y de objetos ordenados por las secuencias. Todo este aparato

7 T. Burge, "Reference and Proper Names".
8 G. Harman, "Moral Relativism Defended".
9 J. Wallace, "Positive, Comparative, Superlative".
10 Véanse los Ensayos 7 y 8.
11 Véanse los Ensayos 6-10 en *Essays on Actions and Events*.
12 Véase el Ensayo 6.
13 M. Dummett, *Frege: Philosophy of Language*.

está considerado correctamente como una construcción teórica, fuera del alcance de la verificación directa. Ha cumplido su tarea sólo bajo la condición de que implique resultados comprobables en forma de oraciones T, y éstas no hacen mención de la maquinaria. Una teoría de la verdad reconcilia de esta manera la exigencia de una teoría que articule la estructura gramática con la exigencia de una teoría que pueda ser probada sólo por lo que dice acerca de las oraciones.

En la obra de Tarski, las oraciones T son consideradas verdaderas pues se presume que la rama derecha del bicondicional es una traducción de la oración para la cual se están dando condiciones de verdad. Pero no podemos presumir de antemano que la traducción correcta puede ser reconocida sin dominar antes la médula de la interpretación radical; para las aplicaciones empíricas, debemos abandonar la presunción. Lo que yo propongo es invertir la dirección de la explicación: presumiendo la traducción, Tarski estaba en condiciones de definir la verdad; la presente idea consiste en tomar a la verdad como básica y extraer una explicación de la traducción o interpretación. Las ventajas, desde el punto de vista de la interpretación radical, son obvias. La verdad es una propiedad única que se fija o no se fija a las emisiones, mientras que cada emisión tiene su propia interpretación; y la verdad tiene más aptitud para conectarse con actitudes más bien simples de los hablantes.

No hay dificultad en reformular la Convención T sin apelar al concepto de traducción: una teoría aceptable de la verdad debe implicar, para cada oración *s* del objeto lenguaje, una oración de la forma: *s* es verdadera si y sólo si *p*, donde "*p*" se reemplaza por cualquier oración que sea verdadera si y sólo si *s* lo es. Dada esta formulación, la teoría se pone a prueba mediante la evidencia de que las oraciones T son simplemente verdaderas; hemos abandonado la idea de que también debemos decir si lo que reemplaza a "*p*" traduce a *s*. Podría parecer que no hay ninguna perspectiva de que exigiendo tan poco de las oraciones T emerja una teoría de la interpretación. Y desde luego que esto sería así si tomáramos a las oraciones T aisladamente. Pero la esperanza es que si aplicamos las restricciones formales y empíricas apropiadas sobre la teoría como un todo, las oraciones T individuales servirán efectivamente para producir interpretaciones.[14]

Todavía debemos decir qué evidencia está disponible para un intérprete; evidencia, vemos ahora, de que las oraciones T son verdaderas. La evidencia no puede consistir en descripciones detalladas de las creencias e intenciones del hablante. Dado que las atribuciones de actitudes, al menos cuando se requiere sutileza, exigen una teoría que deba descansar en una evidencia muy parecida a la de la interpretación. La interdependencia de creencia y significado resulta evidente de esta manera: un hablante sostiene que una oración es verdadera debido a lo que la oración (en su lenguaje) significa, y debido a lo que él cree. Sabiendo

[14] Para requisitos esenciales, véase la nota 11 del Ensayo 2.

que él toma a la oración por verdadera, y conociendo el significado, podemos inferir su creencia; dada suficiente información acerca de sus creencias, podemos tal vez inferir el significado. Pero la interpretación radical debe descansar en evidencia que no presuma un conocimiento de significados o un conocimiento detallado de creencias.

Un buen punto de partida es la actitud de tomar a una oración por verdadera, de aceptarla como verdadera. Esto es, desde luego, una creencia, pero es una actitud única aplicable a todas las oraciones, y por lo tanto no requiere de nosotros que seamos capaces de hacer distinciones finamente discriminadas entre creencias. Es plausible considerar que un intérprete es capaz de identificar esta actitud antes de poder interpretar, dado que puede saber que una persona pretende expresar una verdad al emitir una oración sin tener la menor idea de *qué* verdad se trata. No es que la aserción sincera sea la única razón para suponer que una persona considera una oración como verdadera. Mentiras, órdenes, historias, ironía, si son detectadas como actitudes, pueden revelar si un hablante considera sus oraciones como verdaderas. No hay razón para descartar otras actitudes para con las oraciones, tales como desear su verdad, querer probar su verdad, creer que uno va a probar su verdad, y así por el estilo, pero me inclino a pensar que toda evidencia de esta clase puede agruparse en términos de considerar a las oraciones como verdaderas.

Supóngase, entonces, que la evidencia disponible consiste justamente en que los hablantes del lenguaje a ser interpretado sostienen que diversas oraciones son verdaderas en ciertos tiempos y bajo circunstancias especificadas. ¿Cómo puede emplearse esta evidencia para respaldar una teoría de la verdad? Por un lado, tenemos oraciones T, de la forma:

> (T) *"Es regnet"* es verdadero-en-alemán cuando está hablado por x en el tiempo t si y sólo si está lloviendo cerca de x en t.

Por el otro lado, tenemos la evidencia, de la forma:

> (E) Kurt pertenece a la comunidad de habla alemana y Kurt sostiene que es verdadero *"Es regnet"* el sábado al mediodía y está lloviendo cerca de Kurt el sábado al mediodía.

Pienso que deberíamos considerar (E) como evidencia de que (T) es verdadera. Puesto que (T) es un condicional universalmente cuantificado, el primer paso sería reunir más evidencia para respaldar la afirmación de que:

> (GE) (x) (t) [si x pertenece a la comunidad de habla alemana luego (x sostiene que es verdadero *"Es regnet"* en t si y sólo si está lloviendo cerca de x en t)].

La apelación a una comunidad de habla simplifica la tarea sin supo-

ner nada establecido: los hablantes pertenecen a la misma comunidad de habla si las mismas teorías de interpretación dan resultados para ellos.

La objeción obvia es que Kurt, o cualquier otra persona, puede estar equivocado respecto de si está lloviendo cerca de él. Y esto es desde ya una razón para no tomar (E) como evidencia concluyente para (GE) o para (T); y una razón para no pretender que generalizaciones como (GE) sean más que generalmente verdaderas. El método consiste más bien en obtener el mejor ajuste. Queremos una teoría que satisfaga las restricciones formales de una teoría de la verdad, y que maximice el acuerdo, en el sentido de hacer que Kurt (y otros) estén en lo cierto, hasta donde nosotros podamos discernirlo, tan a menudo como sea posible. El concepto de maximización no puede tomarse aquí literalmente, puesto que las oraciones son infinitas en número, y en todo caso una vez que la teoría comienza a tomar forma cobra sentido aceptar el error inteligible y admitir la relativa probabilidad de diversas clases de error.[15]

El proceso de conformación de una teoría de la verdad para una lengua nativa desconocida podría a grandes rasgos seguir los siguientes pasos. Primero buscamos la mejor manera de ajustar nuestra lógica, hasta el punto requerido para obtener una teoría que satisfaga la Convención T, en el nuevo lenguaje; esto puede significar leer la estructura lógica de la teoría de cuantificación de primer orden (más la identidad) dentro del lenguaje, no tomando las constantes lógicas una por una, sino tratando a esta gran porción de lógica como a una red que se instalará sobre el lenguaje en una única operación relámpago. La evidencia consiste aquí en clases de oraciones que siempre se consideran verdaderas o siempre se consideran falsas por casi todas las personas casi todo el tiempo (verdades lógicas potenciales) y en patrones de inferencia. El primer paso identifica predicados, términos singulares, cuantificadores, conectivos e identidad; en teoría, establece cuestiones de forma lógica. El segundo paso se concentra en oraciones con deícticos; aquellas oraciones que a veces consideradas verdaderas y a veces falsas de acuerdo con cambios descubribles en el mundo. Este paso en conjunción con el primero limita las posibilidades de interpretar predicados individuales. El último paso se ocupa de las oraciones restantes, aquellas para las cuales no hay un acuerdo uniforme, o cuyo valor de verdad considerado no depende sistemáticamente de cambios en el medio ambiente.[16]

[15] Para más detalles sobre la obtención de un "mejor ajuste", véanse los Ensayos 10-12.

[16] Los lectores que aprecien en qué grado esta consideración es paralela a la consideración de la traducción radical de Quine del Capítulo 2 de *Word and Objet* apreciarán también las diferencias: la restricción semántica de mi método fuerza una estructura cuantificacional sobre el lenguaje a ser interpretado, lo cual probablemente no deja espacio para una indeterminación de forma lógica; la noción de significado-estímulo no desempeña función alguna en mi método, pero su lugar lo ocupa la referencia a las características objetivas del mundo que se ven alteradas en conjunción con los cambios de actitud hacia la verdad de las oraciones; el principio de caridad, al cual Quine enfatiza sólo en conexión con la identificación de los conectivos oracionales (puros), es aplicado por mí desde el principio hasta el final.

Este método pretende resolver el problema de las interdependencia de creencia y significado tomando a la creencia como constante hasta donde sea posible en tanto se resuelve el significado. Esto se consigue mediante la asignación de condiciones de verdad a oraciones extranjeras que hacen que los hablantes nativos tengan razón cuando es plausiblemente posible, por supuesto según nuestro propio punto de vista de lo que es correcto. El procedimiento lo justifica el hecho de que tanto el desacuerdo como el acuerdo son inteligibles sólo dentro de un marco de amplio acuerdo. Aplicado al lenguaje, este principio significa: cuantas más oraciones convoquemos para aceptar o rechazar (a través o no de un medio de interpretación), mejor comprenderemos al resto, nos pongamos o no de acuerdo respecto de ellas.

El consejo metodológico de interpretar en una forma que optimice el acuerdo no debe concebirse apoyado en una presunción caritativa acerca de la inteligencia humana que podría resultar falsa. Si no podemos encontrar una forma de interpretar que las emisiones y otras manifestaciones de la conducta de una creatura son reveladoras de un conjunto de creencias ampliamente consistentes y verdaderas según nuestros propios patrones, no tenemos razones para considerar que esa creatura es racional, tiene creencias o dice algo.

Aquí me gustaría insertar una observación acerca de la metodología de mi propuesta. En filosofía estamos acostumbrados a definiciones, análisis, reducciones. Por lo general su intención es llevarnos de conceptos mejor comprendidos, o claros, o más básicos epistemológica u ontológicamente, a otros que queremos comprender. El método que he sugerido no encaja en ninguna de estas categorías. He propuesto una relación más flexible entre los conceptos a ser iluminados y los relativamente más básicos. En el centro campea una teoría fomal, una teoría de la verdad, la cual impone una estructura compleja sobre las oraciones que contienen las nociones primitivas de verdad y satisfacción. Estas nociones reciben aplicación por medio de la forma de la teoría y de la naturaleza de la evidencia. El resultado es una teoría parcialmente interpretada. La ventaja del método descansa no en su recurso libre a la noción de apoyo evidencial sino en la idea de una teoría poderosa interpretada en el punto más ventajoso. Esto nos permite reconciliar la necesidad de una estructura semánticamente articulada con una teoría comprobable solamente a nivel oracional. El beneficio más sutil está en que una evidencia muy ligera en favor de cada uno de los infinitos puntos potenciales puede proporcionar ricos resultados, incluso con respecto a los puntos. Conociendo tan sólo las condiciones bajo las cuales los hablantes consideran verdaderas a las oraciones, podemos extraer, dada una teoría satisfactoria, una interpretación de cada oración. Aún resta probar esta última afirmación. La teoría misma en el mejor de los casos da condiciones de verdad. Lo que necesitamos demostrar es que si una teoría así satisface las restricciones que hemos especificado, puede ser usada para producir interpretaciones.

3. *Si sabemos que una teoría de la verdad satisface los criterios formales y empíricos descritos, ¿podemos interpretar emisiones del lenguaje para el cual ella es una teoría?*

Una teoría del lenguaje implica una oración T para cada oración del lenguaje, y una oración T da condiciones de verdad. Resulta tentador, por ello, decir simplemente que una oración T "da el significado" de una oración. Por supuesto, no diríamos que lo hace nombrando o describiendo una entidad que es un significado, sino simplemente diciendo bajo qué condiciones una emisión de la oración es verdadera.

Pero tras reflexionar resulta claro que una oración T no da el significado de la oración a la cual concierne: La oración T sí fija el valor de verdad relativo a ciertas condiciones, pero no dice que la oración del lenguaje objeto sea verdadera *porque* las condiciones son válidas. Empero, si se hiciera tanto caso a los valores de verdad, la oración T para "La nieve es blanca" bien podría decir que ésta es verdadera si y sólo si el césped es verde o 2+2=4, como podría decir que es verdadera si y sólo si la nieve es blanca. Podemos tener confianza, quizá, en que ninguna teoría satisfactoria de la verdad producirá oraciones T tan anómalas, pero esta confianza no nos da autorización para sacar más provecho de las oraciones T.

Hay una jugada que podría parecer de utilidad y que consiste en afirmar que no es la oración T sola, sino la prueba canónica de una oración T, la que nos permite interpretar la oración extraña. Dada una teoría de la verdad, una prueba canónica es fácil de construir, moviéndose como lo hace a través de una hilera de bicondicionales, y requiriendo para su singularidad solamente decisiones ocasionales para gobernar la precedencia de izquierda o de derecha. La prueba efectivamente refleja la forma lógica que la teoría asigna a la oración, y por esto podría pensarse que revela algo acerca del significado. Pero de hecho no sabríamos más que antes acerca de cómo interpretar si todo lo que supiéramos fuera que una determinada secuencia de oraciones es la prueba, para alguna teoría verdadera, de una oración T particular.

Una sugerencia final en la misma línea de pensamiento sería que podemos interpretar una oración particular siempre que conozcamos una teoría correcta de la verdad que trate del lenguaje de la oración. Pues entonces conocemos no solamente la oración T para la oración a ser interpretada, sino que también "conocemos" las oraciones T para todas las demás oraciones; y, por supuesto, todas las pruebas. Luego veríamos el papel de la oración en el lenguaje como un todo, conoceríamos el rol de cada parte significativa de la oración, y sabríamos acerca de las conexiones lógicas entre esta oración y otras.

Si supiéramos que una oración T satisface la Convención T de Tarski, sabríamos que es verdadera, y podríamos usarla para interpretar una oración porque sabríamos que la rama derecha del bincondicional traduce la oración a ser interpretada. Nuestro problema actual emana del hecho de que en la interpretación radical no podemos suponer que una oración T satisface el criterio de traducción. Hemos estado de-

jando de lado, sin embargo, que ya proporcionamos un criterio alternativo: este criterio consiste en que la totalidad de las oraciones T deberían ajustarse óptimamente (en el sentido descrito más atrás) a la evidencia acerca de las oraciones consideradas verdaderas por los hablantes nativos. La presente idea es que lo que Tarski daba directamente por supuesto para cada oración T puede ser extraído indirectamente mediante una restricción holística. Si esta restricción es adecuada, cada oración T producirá en efecto una interpretación aceptable.

Una oración T de una teoría empírica de la verdad puede entonces emplearse para interpretar una oración siempre que conozcamos también la teoría que ella implica, y que sepamos que es una teoría que satisface los criterios formales y empíricos.[17] Pues si las restricciones son adecuadas, el rango de teorías aceptables será tal que cualquiera de ellas proporcionará alguna interpretación correcta para cada emisión potencial. Para ver cómo funcionaría esto, aceptemos por un momento la hipótesis absurda de que las restricciones reducen las teorías posibles a una, y ella implica la oración T (T) previamente discutida. Entonces nos vemos justificados al usar esta oración T para interpretar la emisión de Kurt de *"Es regnet"* como su dicho de que está lloviendo. No parece probable, dada la naturaleza flexible de las restricciones, que todas las teorías aceptables vayan a ser idénticas. Cuando se haya incorporado toda la evidencia, todavía quedarán, tal como lo ha enfatizado Quine, los compromisos entre las creencias que nosotros atribuimos a un hablante y las interpretaciones que damos a sus palabras. Pero la indeterminación resultante no puede ser tan grande, a no ser porque toda teoría que supere las pruebas servirá para producir interpretaciones.

[17] Véase la nota 11 del Ensayo 2 y el Ensayo 12.

10

La creencia y el fundamento del significado

Significado y creencia desempeñan papeles interrelacionados y complementarios en la interpretación del discurso. Al poner énfasis en la conexión entre nuestras razones para atribuir creencias a los hablantes, y nuestras razones para asignar significados a sus emisiones, espero explicar algunos rasgos problemáticos tanto de la creencia como del significado.

Estamos interpretando un fragmento de conducta lingüística cuando decimos lo que las palabras de un hablante significan en una ocasión de uso. La tarea puede verse como una redescripción. Sabemos que las palabras *"Es schneit"* han sido emitidas en una ocasión particular y queremos redescribir esta emisión como un acto de decir que está nevando.[1] ¿Qué necesitamos saber para estar en condiciones de redescribir el habla de esta manera, es decir, para interpretar las emisiones de un hablante? Puesto que un intérprete competente puede interpretar cualquiera de las infinitas emisiones potenciales (o bien podríamos decir lo mismo nosotros), no podemos especificar lo que él conoce mediante un listado de casos. Sabe, por ejemplo, que al emitir *"Es schneit"* bajo ciertas condiciones y con una determinada intención, Karl ha dicho que está nevando; pero hay una cantidad interminable de otros casos. Lo que debemos hacer entonces es enunciar una teoría finita a partir de la cual derivan las interpretaciones particulares. La teoría puede usarse para describir un aspecto de la competencia del intérprete en la comprensión de lo que se dice. También podemos sostener, si lo deseamos, que hay un mecanismo en el intérprete que corresponde a la teoría. Si esto significa solamente que hay alguno u otro mecanismo que ejecuta esa tarea, resulta difícil ver cómo la afirmación podría dejar de ser verdadera.

La teoría de la interpretación es un trabajo común del lingüista, del psicólogo y del filósofo. El sujeto que trata es la conducta de un hablante o hablantes, y nos dice qué significan determinadas emisiones de aquéllos. Por último, la teoría puede usarse para describir lo que cada intérprete sabe, esto es, un subconjunto infinito y especificable de las

[1] Uso de la expresion "dice que" en el presente contexto de tal manera que un hablante dice (en una ocasión particular) que está nevando si y sólo si emite palabras que (en esa ocasión) significan que está nevando. Por lo tanto un hablante puede decir que está nevando si *él* significar, o afirmar, que está nevando.

verdades de la teoría. A continuación diré un poco, y supondré mucho, acerca de la forma que puede tomar una teoría de la interpretación. Pero quiero poner el acento en la cuestión de cómo darse cuenta de que una teoría de este tipo es verdadera.

Hay una respuesta que salta a la vista. La teoría es verdadera si sus implicaciones empíricas son verdaderas; podemos verificar la teoría tomando muestras de sus implicaciones en busca de verdad. En el caso presente, esto significa notar si las interpretaciones típicas que una teoría produce para las emisiones de un hablante son correctas o no. Acordamos que todo intérprete competente sabe si las implicaciones relevantes son verdaderas; por lo tanto todo intérprete competente puede verificar una teoría de esta forma. Esto no significa, desde luego, que encontrar una teoría verdadera sea trivial; sí significa que dada una teoría, su puesta a prueba puede no necesitar requisitos misteriosos.

La pregunta original, empero, es cómo sabemos que una interpretación particular es correcta, y nuestra respuesta inmediata no está orientada hacia esta cuestión. No cabe duda de que una emisión puede interpretarse mediante una teoría correcta, pero si el problema es determinar cuándo una interpretación es correcta, no sirve de nada respaldar la teoría que le da origen dando ejemplo de interpretaciones correctas. En apariencia nos enfrentamos a un *impasse*: necesitamos la teoría antes de poder reconocer evidencia a su favor.

El problema es grave pues las emisiones ininterpretadas parecen ser la base evidencial apropiada para una teoría del significado. Si una teoría aceptable pudiera respaldarse mediante tal evidencia, eso constituiría un progreso conceptual, pues la teoría sería de naturaleza específicamente semántica mientras que la evidencia estaría descrita en términos no semánticos. Un intento de acumular evidencia aun más elemental, digamos evidencia conductista, sólo podría complicar la tarea de construcción de la teoría, si bien podría tornarla más satisfactoria. En todo caso podemos acometer la empresa menor sin avergonzarnos.

Una fuente central de problemas es la forma en que creencias y significados conspiran contra la explicación de las emisiones. Un hablante que considera que una oración es verdadera en una ocasión lo hace en parte por lo que él significa, o significaría, mediante una emisión de esa oración, y en parte por lo que él cree. Si sólo podemos proseguir en base a emisión honesta, no podemos inferir la creencia sin conocer el significado, y no tenemos esperanza de inferir el significado sin la creencia.

Hay varias estrategias para romper este círculo que se postulan por sí mismas. Una es encontrar evidencia de lo que significan las palabras que sea independiente de la creencia. Tendría que ser independiente de las intenciones, deseos, reproches, apetencias, consentimientos y también convenciones, pues todos ellos tienen una componente de creencia. Quizás haya quien piense que sería posible establecer la corrección de una teoría de la interpretación sin conocer o establecer mucho acerca de las creencias, pero no es fácil imaginar cómo podría lograrse esto.

Mucho más plausible es la idea de derivar una teoría de la interpretación a partir de información detallada acerca de las intenciones, deseos y creencias de los hablantes (o intérpretes, o ambos). Considero que ésta es la estrategia de aquellos que se abocan a definir o explicar el significado lingüístico sobre la base de intenciones, usos, propósitos, funciones y demás, de índole no lingüística: son éstas las tradiciones de Mead y Dewey, de Wittgenstein y Grice. Yo pienso que esta estrategia tampoco satisfaría las necesidades presentes.

Puede no haber nada erróneo, desde luego, en la máxima metodológica que dice que cuando surgen problemas desconcertantes acerca de significados, referencia, sinonimia y demás, debemos recordar que estos conceptos, al igual que aquellos mismos de palabra, oración y lenguaje, nos abstraen de las transacciones y entornos sociales que les dan el contenido que tienen. Los conceptos lingüísticos y semánticos de todos los días son parte de una teoría intuitiva de organización de datos más primitivos, por lo que sólo puede resultar más confusión de un tratamiento de estos conceptos y de sus objetos supuestos como si tuvieran vida propia. Pero esta observación no puede responder a la cuestión de cómo sabemos cuando una interpretación de una emisión es correcta. Si nuestros conceptos ordinarios sugieren una teoría confusa, debemos buscar una teoría mejor, no dejar de lado la teorización. Puede no encontrarse objeción en detallar las complicadas e importantes relaciones entre lo que las palabras de un hablante significan y sus intenciones y creencias no lingüísticas. Tengo mis dudas respecto de la posibilidad de *definir* el significado lingüístico en términos de intenciones y creencias no lingüísticas, pero esas dudas, si no las fuentes de esas dudas, son irrelevantes para el tema presente.

El tema presente es la naturaleza de la evidencia para la adecuación de una teoría de la interpretación. La evidencia debe ser descriptible en términos no semánticos o no lingüísticos si es que va a responder a la pregunta que hemos formulado; debe ser también evidencia que podamos imaginar al alcance del investigador virgen antes de que éste ya esté en posesión de la teoría de la cual se supone que ella es evidencia. Es aquí donde yo avizoro problemas. Se presenta un obstáculo de principios, y no meramente práctico, en la verificación de la existencia de creencias e intenciones detalladas, generales y abstractas, cuando se es incapaz de determinar lo que significan las palabras del hablante. Percibimos lo suficientemente bien el absurdo de tratar de comprender sin preguntarle a él si es que alguien cree que hay un objetivo mayor, o si es que él pretende, al hacer determinados ruidos, que alguien deje de fumar al reconocer ese alguien que los ruidos se hacían con esa intención. El absurdo descansa no en el hecho de que sería muy difícil descubrir estas cosas sin lenguaje, sino en el hecho de que no tenemos una buena idea de cómo comenzar a autenticar la existencia de tales actitudes cuando la comunicacion no es posible.

Esta cuestión no queda expuesta de manera feliz diciendo que nuestras complejas creencias, intenciones y pensamientos son como emisiones silenciosas. Mi afirmación consiste solamente en que dar un

sentido detallado a las intenciones y creencias de una persona no puede ser independiente de dar sentido a sus emisiones. Si esto es así, luego un inventario de las creencias e intenciones complejas de un hablante no puede ser la evidencia de la verdad de una teoría para interpretar su conducta de habla.

Dado que no podemos tener esperanzas de interpretar la actividad lingüística sin conocer lo que un hablante cree, y que no podemos fundar una teoría de lo que él significa sobre un descubrimiento previo de sus creencias e intenciones, concluyo que al interpretar emisiones desde los comienzos —en la interpretación *radical*— de alguna manera debemos proveer simultáneamente una teoría de la creencia y una teoría del significado. ¿Cómo es posible esto?

De manera de hacer que el problema sea lo suficientemente definido y simple como para una discusión relativamente breve, permítaseme introducir un cambio en la descripción de la base evidencial para una teoría de la interpretación. En vez de emisiones de expresiones, quiero considerar una cierta actitud hacia las expresiones, una actitud que puede o no revelarse en las emisiones reales. Esta actitud es la de considerar verdadero, relativizado al tiempo. También podemos suponer que disponemos de todo lo que pueda saberse sobre tales actitudes, pasadas, presentes y futuras. Por último, quiero imaginar que podemos describir las circunstancias externas bajo las cuales las actitudes son válidas o no lo son. Un ejemplo típico de la clase de evidencia disponible sería entonces el siguiente: un hablante considera que *"Es schneit"* es verdadero cuando y sólo cuando está nevando. Espero que se concederá que es plausible decir que podemos darnos cuenta cuándo un hablante considera que una oración es verdadera sin conocer lo que él significa con la oración, o qué creencias sostiene respecto de la materia desconocida de que ella trata, o qué intenciones detalladas lo impulsan o lo impulsarían a emitirla. A menudo se argumenta que debemos presumir que la mayor parte de las emisiones de un hablante corresponden a oraciones que él considera verdaderas: si esto es correcto, la disponibilidad independiente de la base evidencial está asegurada. Pero presunciones más débiles también funcionarían, puesto que incluso el mentiroso compulsivo y el eterno bromista pueden ser descubiertos.

El problema, entonces, es el siguiente: suponemos que conocemos qué oraciones un hablante considera verdaderas, y cuándo, y queremos conocer lo que significa y cree. Tal vez podríamos resolver el caso si conociéramos lo suficiente acerca de sus creencias e intenciones, pero esto no es posible sin el previo acceso a una teoría de la interpretación. Dadas las interpretaciones, podríamos extraer las creencias de la base evidencial, pero esto supone lo que queremos conocer.

Me ha impactado la analogía con un conocido problema de teoría de la decisión. Supóngase que un agente se muestra indiferente entre obtener $ 5,00, y una apuesta que le ofrece $ 11,00 si una moneda que se arroja cae cara hacia arriba, y $ 0,00 si cae cruz. Podríamos explicar (digamos, "interpretar") su indiferencia suponiendo que el dinero tiene para él una utilidad marginal en disminución: $ 5,00 está a mitad de ca-

mino en su escala de valores subjetiva entre $ 0,00 y $ 11,00. Llegamos a esto al presumir que la apuesta vale la suma de los valores de los posibles resultados moderada por sus probabilidades. En este caso, presumimos que cara y cruz son igualmente probables. Por desgracia, hay una explicación alternativa igualmente admisible: dado que obviamente $ 5,00 no está a mitad de camino entre $ 0,00 y $ 11,00, el agente debe creer que es más probable que salga cruz que cara; si hubiera pensado que cara y cruz tienen igual probabilidad, ciertamente habría preferido la apuesta, que equivaldría entonces a una oferta directa de $ 5,50.

La cuestión es obvia. Las elecciones entre apuestas son resultado de dos factores psicológicos, los valores relativos que el elector adjudica a los resultados, y la probabilidad que asigna a esos resultados, condicional en su elección. Dadas las creencias del agente (sus probabilidades subjetivas) es fácil computar sus valores relativos a partir de sus elecciones; dado sus valores, podemos inferir sus creencias. Pero dadas solamente sus creencias, ¿cómo podemos obtener sus creencias y sus valores?

El problema es muy parecido al problema de la interpretación. La solución en el caso de la decisión es clara y satisfactoria; no se dispone de nada así en la teoría del significado. Pero pienso que uno todavía puede ver la posibilidad de aplicar una estrategia análoga. La propuesta de Frank Ramsey para enfrentar el problema de la teoría de la decisión, algo simplificado, es la siguiente.[2] Supóngase que hay dos alternativas, obtener $ 11,00 y obtener $ 0,00, y que hay un suceso *E* tal que el agente es indiferente entre las dos apuestas siguientes: Apuesta Uno: si *E* sucede el agente recibe $ 11,00; si *E* no sucede recibe $ 0,00. Apuesta Dos: si *E* sucede obtiene $ 0,00; si *E* no sucede obtiene $ 11,00. La indiferencia del agente hacia las apuestas muestra que él debe juzgar que es tan probable que *E* ocurra como que no. Pues si él pensara que es más probable que *E* ocurra, preferiría la primera apuesta que le promete $ 11,00 si *E* ocurre, y si pensara que *E* es más probable que no ocurra preferiría la segunda apuesta que premia el no acaecimiento de *E* con $ 11.00. Esto resuelve, para la teoría de la decisión, el problema de cómo separar probabilidad subjetiva de utilidad subjetiva, pues una vez que un suceso como *E* se descubre, es posible confeccionar una escala de otros valores, y luego determinar las probabilidades subjetivas de todos los sucesos.

En esta versión de la teoría de la decisión, la base evidencial consiste en preferencias entre alternativas, algunas de ellas apuestas; la preferencia corresponde aquí a la actitud de considerar verdadero para el caso de la interpretación, tal como yo presento el problema. Las elecciones reales en la teoría de la decisión corresponden a emisiones reales en la interpretación. La explicación de una preferencia particular involucra la asignación de una clasificación comparativa de valores y una

[2] F. P. Ramsey, "Truth and Probability".

evaluación de probabilidades. El hecho de respaldar esta explicación no proviene de un nuevo tipo de visión de las actitudes y creencias del agente, sino de un mayor número de observaciones de preferencias del mismo tipo a ser explicado. En pocas palabras, para explicar (digamos, interpretar) una elección o preferencia particular, observamos *otras* elecciones o preferencias; éstas respaldarán una teoría sobre cuya base puede explicarse la elección o preferencia original. Las atribuciones de valores y probabilidades subjetivas forman parte de la estructura teórica, y son formas convenientes de resumir hechos acerca de la estructura de las preferencias básicas; y no hay modo de probarlos independientemente. En líneas generales, mi idea es que debemos ver a los significados y creencias como constructos interrelacionados de una única teoría tal como ya vemos a los valores y probabilidades subjetivas como constructos interrelacionados de la teoría de la decisión.

Una forma de representar algunos de los hechos explicativos acerca de la conducta de elección producidos por un teoría de la decisión consiste en asignar números para medir, digamos, los valores subjetivos de los resultados para un agente particular. Así podríamos asignar los números 0, 1 y 2 como medidas de los valores para alguien que los aplica a la obtención de $ 0,00, $ 5,00 y $ 11,00 respectivamente. Esto podría sugerir a un incauto que para ese agente $ 11,00 vale el doble que $ 5,00. Sólo mediante el estudio de la teoría subyacente emergería la verdad de que la asignación de números para mediar las utilidades es única en cuanto refiere una transformación lineal, pero no va más allá de eso. Los números 2, 4 y 6 podrían haber funcionado igualmente bien para el registro de los hechos, pero 6 no es el doble de 4. La teoría refleja comparaciones de diferencias, pero no comparaciones de magnitudes absolutas. Cuando representamos los hechos de preferencia, utilidad y probabilidad subjetiva mediante la asignación de números, sólo estamos usando algunas de las propiedades de los números para captar un patrón justificado empíricamente. Otras propiedades de los números empleados pueden entonces elegirse arbitrariamente, como el punto cero y la unidad para medir utilidad o temperatura.

Los mismos hechos pueden ser representados mediante asignaciones muy diferentes de números. En la interpretación del habla, la introducción de supuestas entidades como las proposiciones entendidas como significados de oraciones u objetos de creencia pueden llevarnos al error de pensar que la evidencia justifica, o tendría que justificar, un tipo de unicidad o singularidad que no justifica. En el caso de la teoría de la decisión, podemos establecer exactamente cuáles propiedades de los números son relevantes para la medición de la utilidad y cuáles para la medición de la probabilidad. Siendo las proposiciones mucho más vagas que los números, no resulta claro hasta qué punto están sobreexigidas en su tarea.

No hay tan sólo una simple analogía entre la teoría de la decisión y la teoría de la interpretación, hay también una conexión. Visto desde el lado de la teoría de la decisión, existe lo que Ward Edwards llamó una vez "problema de presentación" para las aplicaciones empíricas de la

teoría de la decisión. Para entender las preferencias de un agente, particularmente entre las apuestas complejas, es obviamente necesario describir las opciones en palabras. Pero ¿cómo puede el experimentador conocer lo que esas palabras significan para el sujeto? El problema no es meramente teórico: es bien sabido que dos descripciones de lo que el experimentador toma como la misma opción pueden generar respuestas muy diferentes en el sujeto. Enfrentamos un problema que discutimos hace un momento en conexión con la interpretación: no es razonable suponer que podemos interpretar la conducta verbal sin información en detalle acerca de creencias e intenciones, ni tampoco es razonable imaginar que podemos justificar la atribución de preferencias entre opciones complejas a menos que podamos interpretar la conducta de habla. Una teoría radical de la decisión debe incluir una teoría de la interpretación y no puede presuponerla.

Visto desde el lado de una teoría de la interpretación, se presenta la dificultad obvia de determinar cuándo una persona acepta una oración como verdadera. La teoría de la decisión, y las ideas de sentido común que la respaldan, ayudan a reforzar la visión de que las creencias se comprenden mejor en su función de racionalizar elecciones o preferencias. Aquí estamos considerando solamente una clase especial de creencia, la creencia de que una oración es verdadera. Pero aun en este caso, sería mejor si pudiéramos llegar, a través de la creencia, a una preferencia que podría mostrarse en una elección. No dispongo de una propuesta detallada para indicar en este momento cómo podría o debería hacerse esto. Richard Jeffrey ha dado un primer paso importante.[3] El elimina algunas confusiones problemáticas de la teoría de Ramsey reduciendo la ontología más bien oscura de esta teoría, que trataba con sucesos, opciones y proposiciones, a una ontología de proposiciones solamente. Así las preferencias entre proposiciones consideradas verdaderas pasan a ser la base evidencial, de tal manera que la teoría revisada nos permite hablar de grados de creencia en la verdad de la proposiciones, y de la fuerza relativa de los deseos de que las proposiciones sean verdaderas. Tal como lo señala Jeffrey, para los propósitos de esta teoría los objetos de estas diversas actitudes también pueden tomarse como oraciones. Si se realiza este cambio podemos unificar los sujetos de la teoría de la decisión y de la teoría de la interpretación. Jeffrey supone, por supuesto, que los agentes y los constructores de la teoría entienden las oraciones de la misma manera. Pero pueden unirse las dos teorías dejando de lado esta suposición. La teoría para la cual debemos finalmente esforzarnos es aquella que tome como base evidencial las preferencias entre oraciones: preferencias para considerar verdadera a una oración y no a otra. La teoría explicaría luego las preferencias individuales de este tipo atribuyendo creencias y valores al agente, y significados a sus palabras.[4]

[3] R. Jeffrey, *The Logic of Decision*.

[4] Respecto del progreso en el desarrollo de esa teoría, véase mi "Toward a Unified Theory of Meaning and Action".

No seguiré especulando en este trabajo acerca de las posibilidades de lograr una teoría integrada de la decisión y de la interpretación; vuelvo entonces al problema de interpretar emisiones en base a la información respecto de cuándo, y bajo qué circunstancias externas, las oraciones que ellas ejemplifican son consideradas verdaderas. Las ideas centrales de lo que he dicho hasta ahora pueden resumirse así: los hechos conductales o de disposición que pueden ser descritos en formas que no suponen interpretaciones, pero sobre los cuales puede basarse una teoría de la interpretación, serán necesariamente vectores de significado y de creencia. Un resultado de esto es que para interpretar una emisión particular se necesita construir una amplia teoría para la interpretación de un número potencialmente infinito de emisiones. La evidencia para la interpretación de una emisión particular tendrá que ser entonces una evidencia para la interpretación de todas las emisiones de un hablante o comunidad. Por último, si entidades tales como significados, proposiciones y objetos de creencia ocupan un espacio legítimo en la explicación de la conducta de habla, esto sólo se debe a que puede demostrarse que ellas desempeñan un papel fundamental en la construcción de una teoría adecuada. No hay una razón para creer de antemano que estas entidades nos prestarán alguna ayuda, y por ello la identificación de los significados de expresiones o de los objetos de creencia no puede constituir una meta independiente de una teoría o análisis.

La apreciación de estas ideas, que debemos mayormente a Quine, representa uno de los pocos saltos verdaderos en el estudio del lenguaje. He puesto las cosas a mi modo, pero pienso que las diferencias entre nosotros son más bien cuestiones de énfasis que de sustancia. Buena parte de los escritos de Quine se concentran —es comprensible— en demoler una confianza equivocada en la utilidad o inteligibilidad de conceptos como analiticidad, sinonimia y significado. Yo he tratado de acentuar lo positivo. Quine, al igual que todos nosotros, quiere proporcionar una teoría de la interpretación. Su animadversión hacia los significados está concebida para desalentar falsos comienzos; pero sus argumentos en apoyo de esa rigurosidad aportan fundamentos para una teoría aceptable.

He aceptado lo que pienso que es esencialmente la visión de Quine del problema de la interpretación, y la estrategia que quiero proponer para su solución obviamente le deberá mucho a él. Habrá también algunas diferencias. Una diferencia concierne a la forma que debería tomar la teoría . Quine nos induciría a confeccionar un manual de traducción (una función, dada recursivamente) que produzca una oración en el lenguaje del intérprete para cada oración del hablante (o más de una oración en caso de ambigüedad). Para interpretar una emisión particular uno daría la oración que la traduce y especificaría el manual de traducción. Además, sería necesario conocer exactamente qué información preservaría un manual de traducción que se atuviera a las restricciones empíricas: esto es, qué sería invariante entre un manual de traducción acectable y otro.

Sugiero hacer la teoría de carácter explícitamente semántico; de

hecho, debería tomar la forma de una teoría de la verdad al estilo de Tarski.[5] Al estilo de Tarski, pero con modificaciones para ajustarse a los problemas presentes. Por un lado, estamos detrás de una *teoría* de la verdad, en tanto que Tarski se interesaba por una definición explícita. No discutiré esta modificación ahora: ella concierne principalmente a la cuestión del grado de riqueza de una ontología disponible en el lenguaje en el cual está dada la teoría. En segundo término, para hacer lugar a la presencia de elementos demostrativos en el lenguaje natural es necesario relativizar la teoría de la verdad a tiempos y hablantes (y posiblemente a algunas otras cosas). La tercera modificación es más seria y llega al corazón del tema de discusión. La Convención T de Tarski exige a una teoría de la verdad que ponga condiciones a algún predicado, como "es verdadero", de tal modo que implique a todas las oraciones de una forma determinada. Estas son precisamente aquellas oraciones con la forma familiar: " 'La nieve es blanca' es verdadero si y sólo si la nieve es blanca". En los lenguajes formalizados de los que habla Tarski, las oraciones T (como podemos llamar a estos teoremas) se conocen por su sintaxis, y esto sigue siendo verdadero aunque el lenguaje objeto y el metalenguaje sean lenguajes diferentes y aunque sustituyamos las comillas por algo más manejable. Pero en la interpretación radical un test sintáctico de la verdad de las oraciones T carecería de valor, dado que tal test presupondría la comprensión del lenguaje objeto que uno pretende conseguir. La razón es simple: el test sintáctico está ideado meramente para formalizar la relación de sinonimia o traducción, y esta relación se considera no problemática en el trabajo de Tarski sobre la verdad. Nuestro punto de vista invierte el de Tarski: nosotros queremos llegar a una comprensión del significado o de la traducción entendiendo un dominio previo de concepto de verdad. Necesitamos, por lo tanto, una forma de juzgar la aceptabilidad de las oraciones T que no sea sintáctica y que no haga uso de los conceptos de traducción, significado o sinonimia, pero que sea tal que las oraciones T aceptables produzcan efectivamente interpretaciones.

Una teoría de la verdad será materialmente adecuada, esto es, determinará correctamente la extensión del predicado de verdad, siempre y cuando implique, para cada oración *s* del lenguaje objeto, un teorema de la forma "*s* es verdadero si y sólo si *p*" donde "s" es reemplazado por una descripción de *s* y "*p*" es reemplazado por una oración que es verdadera si y sólo si *s* lo es. Para propósitos de interpretación, sin embargo, la verdad en una oración T no es suficiente. Una teoría de la verdad producirá interpretaciones sólo si sus oraciones T enuncian condiciones de verdad en términos de lo que puede verse como "dar el significado" de la oraciones del lenguaje objeto. Nuestro problema es encontrar en la teoría restricciones que sean lo suficientemente fuertes como para garantizar que ella pueda emplearse para la interpretación.

[5] A. Tarski, "The Concept of Truth and Action".

Existen restricciones de naturaleza formal que emanan de la exigencia de que la teoría esté finitamente axiomatizada, y de que satisfaga la Convención T (con las modificaciones apropiadas).[6] Si se considera que el metalenguaje contiene teoría de cuantificación ordinaria, será difícil, si no imposible, descubrir en el lenguaje objeto otra cosa que no sea estructuras cuantificacionales estándar. Esto no significa que cualquier cosa pueda encontrarse en el lenguaje objeto a partir de la simple suposición de que se encuentra en el metalenguaje; por ejemplo, la presencia de operadores modales en el metalenguaje no necesariamente conduce a una teoría de la verdad para un lenguaje objeto modal.

Una teoría satisfactoria, según parece, no puede apartarse mucho de las estructuras cuantificacionales estándar o de la semántica usual de éstas. Debemos suponer que la teoría descansará sobre algo muy parecido al tipo de caracterización recursiva de la satisfacción de Tarski, y que describirá oraciones del lenguaje objeto en términos de pautas familiares creadas por cuantificación y referencia cruzada, predicación, conexiones funcionales de verdad, y otras por el estilo. La relación entre estas pautas semánticamente tratables y la gramática de superficie de las oraciones puede ser, desde luego, muy complicada.

El resultado que se busca al aplicar las restricciones formales es, entonces, ajustar el lenguaje objeto como un todo al lecho procusteano de la teoría de la cuantificación. Si bien esto, indudablemente, puede hacerse de varias maneras, es improbable que las diferencias entre teorías aceptables sean grandes en materia de forma lógica. La identificación de los rasgos semánticos de una oración será entonces esencialmente invariante: las teorías correctas concordarán en un todo respecto de la estructura cuantificacional a ser asignada a una oración dada.

Ya establecidas las cuestiones de forma lógica, las constantes lógicas de la teoría de cuantificación (incluida la identidad) serán descubiertas por fuerza en el lenguaje objeto (bien ocultas, probablemente, debajo de la superficie). Todavía queda por interpretar las demás expresiones primitivas. El problema principal es encontrar una forma sistemática de cotejar los predicados del metalenguaje con los predicados primitivos del lenguaje objeto de modo de obtener oraciones T aceptables. Si los predicados del metalenguaje traducen los predicados del lenguaje objeto, es obvio que las cosas saldrán bien; podría bastar con que tuvieran la misma extensión. Pero resultaría extraño a nuestro programa emplear estos conceptos para formular las restricciones: las restricciones deben tratar solamente con oraciones y con la verdad. Aun así, resulta fácil ver cómo las oraciones T para oraciones con rasgos deícticos limitan netamente la elección de predicados interpretativos; por ejemplo, la oración T para *"Das ist weiss"* debe tener una forma parecida a ésta: "Para todos los hablantes del alemán x y para todos los tiempo t *'Das ist weiss'* es verdadero hablado por x en t si y sólo si el objeto se-

[6] Véanse los Ensayos 5 y 9.

ñalado por x en t es blanco". Como lo ha señalado Quine en sus discusiones sobre relatividad ontológica, puede quedar espacio para ontologías alternativas, y entonces también para sistemas alternativos para interpretar los predicados del lenguaje objeto. Creo que el rango de teorías aceptables de la verdad puede reducirse al punto en que todas las teorías aceptables produzcan oraciones T que podamos tratar como portadoras de interpretaciones correctas, mediante la aplicación de restricciones adicionales que resulten razonables y que no se valgan de principios. Pero reservaré los detalles para otra ocasión.

Mucho más debe decirse, obviamente, acerca de las restricciones empíricas a la teoría: las condiciones bajo la cuales puede aceptarse a una oración T como correcta. Nos hemos puesto de acuerdo en que la base evidencial de la teoría consistirá en hechos acerca de las circunstancias bajo las cuales los hablantes consideran que oraciones de su lenguaje son verdaderas. He insistido en que dicha evidencia se mantiene neutral entre significado y creencia y supone a ninguno de ambos. Es menester mostrar ahora que esos datos pueden proveer un test para la aceptabilidad de las oraciones T.

Propongo que tomemos el hecho de que a los hablantes de un lenguaje consideran que una oración es verdadera (bajo circunstancias observadas) como evidencia *prima facie* de que la oración es verdadera bajo esas circunstancias. Por ejemplo, instancias positivas de "Hablantes (del alemán) consideran *'Es schneit'* verdadero cuando, y sólo cuando, está nevando" deberían tomarse para confirmar no solo la generalización, sino también la oración T " *'Es schneit'* es verdadero (en alemán) para un hablante x en un tiempo t si y sólo si está nevando en t (y cerca de x)".

No puede esperarse que toda la evidencia se oriente de la misma forma. Habrá diferencias de hablante a hablante, y de un tiempo a otro para un mismo hablante, con respecto a las circunstancias bajo las cuales una oración es considerada verdadera. La política general, sin embargo, consiste en elegir condiciones de verdad que funcionen lo mejor posible para hacer que los hablantes consideren verdaderas a la oraciones cuando (de acuerdo con la teoría y con la visión de los hechos del constructor de la teoría) esas oraciones son verdaderas. Esa es la política general, a ser modificada en una variedad de formas obvias. Puede habilitarse a los hablantes a diferir más a menudo y más radicalmente con respecto a algunas oraciones que con respecto a otras, y no hay razón para no tomar encuentra las diferencias individuales observadas o inferidas de las que se piensa que son causas de anomalías (según las ve la teoría).[7]

La construcción de la teoría no puede ser una cuestión de decidirse por una oración T apropiada para una oración del lenguaje objeto por vez; debe erigirse una enorme que preserve las restricciones formales

[7] Para más datos sobre tales modificaciones, véase el Ensayo 11, y en particular D. Lewis, "Radical Interpretation".

discutidas más atrás mientras adecua la evidencia tanto como se pueda. Y, desde luego, el hecho de que una teoría no erija a los hablantes en poseedores universales de verdades no representa una adecuación de la teoría; la meta no persigue el absurdo de hacer desaparecer el disenso y el error. La idea es más bien que sólo el acuerdo amplio da el marco dentro del cual pueden interpretarse disputas y errores. La tarea de dar un sentido a las emisiones y conducta de los otros, aun a su conducta más aberrante, nos exige encontrar una gran dosis de razón y de verdad en ellos. Ver un exceso de sinrazón en los otros es simplemente socavar nuestra capacidad de comprender a qué se debe que sean tan poco razonables. Si la vasta dosis de acuerdo sobre temas corrientes que se supone en la comunicación escapa a la atención, esto es porque las verdades compartidas son demasiadas y demasiado insípidas como para ser dignas de mención. Nosotros queremos hablar de lo que es nuevo, sorprendente, o está en disputa.

Una teoría para interpretar las emisiones de un único hablante, que no se basara en otra cosa que en sus actitudes hacia las oraciones, tendría, podemos estar seguros, muchas rivales igualmente elegibles, pues las diferencias de interpretación podrían neutralizarse mediante apropiadas diferencias en las creencias que se atribuyen. Sin embargo, dada una comunidad de hablantes que aparentemente tiene el mismo repertorio lingüístico, el teórico hará lo posible por hallar una única teoría de la interpretación: esto reducirá grandemente su elección práctica de teorías preliminares para cada hablante individual. (En un diálogo prolongado, uno comienza a la fuerza con una teoría socialmente aplicable, y la refina a medida que acumula evidencia peculiar del otro hablante.)

Lo que hace posible una teoría social de la interpretación es que podemos construir una pluralidad de estructuras de creencia privadas: la creencia se construye para llenar el vacío entre oraciones consideradas verdaderas por individuos y oraciones verdaderas (o falsas) según pautas públicas. La creencia es privada no porque sea accesible a una sola persona, sino porque puede ser idiosincrática. Las atribuciones de creencia son tan verificables públicamente como las interpretaciones, dado que se basan en la misma evidencia: si podemos comprender lo que una persona dice, podemos saber lo que cree.

Si se encara la interpretación de la manera que he estado discutiendo, es improbable que se encuentre sólo una teoría satisfactoria. La indeterminación de interpretación resultante es la contraparte de la indeterminación de traducción de Quine. En mi aproximación, el grado de indeterminación será, pienso, menor que lo que Quine contempla: esto se debe en parte a que yo abogo por la adopción del principio de caridad desde el principio hasta el final, y en parte a que la unicidad de la estructura cuantificacional está aparentemente asegurada si se satisface la Convención T. Pero en cualquier caso la cuestión de la indeterminación no es fundamental para los propósitos de este trabajo. La indeterminación de significado o de traducción no representa un fracaso en la captación de distinciones significativas; más bien resalta el hecho de

que ciertas distinciones aparentes no son significativas. Si hay indeterminación, es porque cuando se ha incorporado toda la evidencia, permanecen abiertas formas alernativas de establecer los hechos. Ya se ha señalado una analogía con la teoría de la decisión: si los números 1, 2 y 3 indican las relaciones significativas en valor subjetivo entre tres alternativa, luego los números -17, -2 y +13 también lo hacen. Una indeterminación de este tipo no puede resultar genuinamente preocupante.

Lo importante es que si, como lo he sugerido, creencia y significado están interrelacionados, entonces la idea de que cada creencia tiene un objeto definido, y la idea de que cada palabra y cada oración tienen un significado definido, no pueden invocarse al describir el objetivo de una teoría exitosa. Pues aun si, contrariamente a lo que puede ser razonable esperar, no hubiera indeterminación alguna, entidades como significados y objetos de creencia carecerían de interés independiente. Podríamos, desde luego, inventar esas entidades con la conciencia limpia si estuviéramos seguros de que no hay variantes permisibles de teorías. Pero si supiéramos esto, sabríamos cómo enunciar nuestras teorías sin mención de los objetos.

Las teorías de creencia y de significado pueden no requerir objetos exóticos, pero sí usan conceptos que las colocan aparte de las ciencias físicas y de otras ciencias no psicológicas: conceptos como los de significado y creencia son fundamentalmente no reducibles a conceptos físicos, neurológicos, o incluso conductistas. Sin embargo, esta irreductibilidad no se debe a la indeterminación de significado o de traducción, pues si yo estoy en lo cierto, la indeterminación es importante sólo para llamar la atención respecto de cómo la interpretación del habla debe marchar codo con codo con la interpretación de la acción en general, y por ende con la atribución de deseos y de creencias. Son más bien los métodos que debemos invocar al construir las teorías de creencia y de significado los que aseguran la irreductibilidad de los conceptos esenciales de esas teorías. Cada interpretación y atribución de actitud es una jugada dentro de una teoría holística, una teoría necesariamente gobernada por un interés por la consistencia y coherencia general con la verdad, y es esto lo que separa para siempre a estas teorías de aquellas que describen objetos inanimados, o que describen los objetos como inanimados.[8]

[8] Véase el Ensayo 8 de *Essays on Actions and Events*.

11

Pensamiento y habla

¿Cuál es la conexión entre pensamiento y lenguaje? Es evidente que el habla depende del pensamiento, pues hablar es expresar pensamientos. Esta dependencia se manifiesta es innumerables formas. Alguien que emite la oración "La vela se apagó" como una oración del castellano debe proponerse emitir palabras que son verdaderas si y sólo si una vela que se indica se apagó en el momento de la emisión, y debe creer que al producir los sonidos que produce está emitiendo palabras que son verdaderas sólo bajo esas circunstancias. Estas intenciones y creencias no son capaces de llamar la atención del hablante fluido. Pero si bien normalmente pueden no atraer atención, su ausencia sería suficiente para mostrar que esa persona no estaba hablando español, y la ausencia de cualquier pensamiento análogo mostraría que ni siquiera estaba hablando.

La cuestión está en el otro caso: ¿puede haber pensamiento sin habla? Una primera y natural reacción afirma que puede haber. Allí está la experiencia, familiar y molesta, de no poder encontrar las palabras para expresar las ideas de uno. En ocasiones uno puede decidir que el escritor del editorial ha expuesto una cuestión mucho mejor de lo que uno podría haberlo hecho. Y está el perro de Norman Malcolm que, habiendo perseguido una ardilla por el bosque, ladra hacia el árbol equivocado.[1] No es difícil confiar en el perro en la creencia de que la ardilla está en ese árbol.

Una intuición definida, si bien más débil, se inclina en el otro sentido. Es posible imaginar que el hablante que no puede hallar las palabras correctas no tiene una idea clara. Las atribuciones de intenciones y de creencias a los perros pecan de antropomorfismo. Un conductismo primitivo, desconcertado por la privacidad de los pensamientos no dichos, puede adoptar con comodidad el punto de vista de que pensar es realmente "hablar consigo mismo": discurso silencioso.

Bajo la superficie de estas tendencias opuestas circulan corrientes muy fuertes, aunque túrgidas, lo que puede explicar por qué los filósofos han preferido en mayor parte tomar partido por alguna a presentar sus argumentos. Sea cual fuere la razón, la cuestión de la relación entre pensamiento y habla parece haber sido encarada rara vez en sí misma.

[1] N. Malcolm, "Thoughtless Brutes".

La suposición usual es que uno u otro, habla o pensamiento, son fáciles de comprender por comparación, y por lo tanto el más oscuro (cualquiera que sea) puede iluminarse analizándolo o explicándolo en términos del otro.

Pienso que esta suposición es falsa: ni el lenguaje ni el pensamiento pueden explicarse plenamente en términos del otro, y ninguno de ambos tiene prioridad conceptual. Los dos, por cierto, están relacionados, en el sentido de que cada uno requiere del otro para su comprensión; pero la relación no es tan completa como para que alguno de ambos baste, incluso cuando se lo refuerza razonablemente, para explicar al otro. Para llevar adelante esta idea se necesita ante todo mostrar cómo el pensamiento depende de habla, y es ésta la tesis que quiero refinar y luego defender.

Atribuimos un pensamiento a una creatura cada vez que empleamos asertivamente una oración positiva cuyo verbo principal es psicológico —en castellano, "cree", "sabe", "espera", "desea", "piensa", "teme", son ejemplos— seguida por una oración y precedida por el nombre o descripción de la creatura. (Un "que" puede seguir optativa o necesariamente al verbo.) Algunas de esas oraciones atribuyen estados, otras dan cuenta de sucesos o procesos: "cree", "piensa" y "quiere" señalan estados, mientras que "comenzó a creer", "olvidó", "concluyó", "notó", "comprueba" señalan sucesos o procesos. Las oraciones que pueden usarse para atribuir un pensamiento exhiben lo que a menudo se da en llamar, o se analiza como, intensionalidad semántica, lo cual significa que la atribución puede cambiarse de verdadera a falsa, o de falsa a verdadera, mediante sustituciones en las oraciones contenidas que no alterarían el valor de verdad de la oración aislada.

Yo no doy por sentado que si una creatura tiene un pensamiento, entonces nosotros podemos, con recursos del tipo recién delineado, atribuir correctamente ese pensamiento a ella. Pero los pensamientos así atribuibles al menos constituyen un buen muestrario de la totalidad.

Es dudoso que los diversos tipos de pensamiento puedan reducirse a uno, o aun a unos pocos: deseo, conocimiento, creencia, temor, interés, para nombrar algunos casos importantes, probablemente son lógicamente independientes, al punto tal que ninguno puede ser definido usando los otros, ni siquiera con el agregado de nociones adicionales como verdad y causa. No obstante, la creencia es esencial para todos los tipos de pensamiento. Si alguien está contento porque, o advierte que, o recuerda que, o sabe que, la pistola está cargada, luego debe creer que la pistola está cargada. Hasta para preguntarse si la pistola está cargada, o para especular acerca de la posibilidad de que la pistola esté cargada, hace falta la creencia, por ejemplo, de que una pistola es un arma, de que es un objeto físico más o menos durable, etcétera. Hay buenas razones para no insistir en alguna lista particular de creencias que sean necesarias en caso de que una creatura vaya a preguntarse si una pistola está cargada. Empero, es necesario que haya innumerables creencias entrelazadas. El sistema de estas creencias identifica un pensamiento localizándolo en un espacio lógico y epistémico.

Para que haya un pensamiento tiene que haber un encuadre de creencias, pero el tener un pensamiento particular no depende del estado de creencia con respecto a ese mismo pensamiento. Si yo considero ir a un determinado comercio, sé que me veré sometido a un grado de molestia y de gasto, y tengo creencias más complicadas acerca del gozo que experimentaré. Disfrutaré escuchando, digamos, las *Grosse Fuge* de Beethoven, pero siempre que la ejecución alcance una calidad razonable, y yo sea capaz de permanecer atento. Tengo el pensamiento de ir al concierto, pero hasta que decida si voy, no tengo una creencia fija de que iré; hasta ese momento, meramente consideraré el pensamiento.

Podemos decir, resumiendo los dos últimos párrafos, que un pensamiento está definido por un sistema de creencias, pero en sí mismo es autónomo respecto de la creencia.

Por lo general pensamos que tener un lenguaje consiste mayormente en ser capaz de hablar, pero en lo que sigue el habla sólo jugará un papel indirecto. En mi argumento lo esencial es la idea de un intérprete, alguien que comprende las emisiones de otro. Las consideraciones a presentar implican, pienso yo, que un hablante debe ser en sí mismo un intérprete de otros, pero no trataré de demostrar que un intérprete debe ser un hablante, si bien puede haber buenas razones para sostenerlo. Tal vez valga la pena señalar que la noción de un lenguaje, o de dos personas hablando el mismo lenguaje, no parece necesitarse aquí. Dos hablantes podrían interpretar sus emisiones recíprocamente sin que hubiera, en un sentido habitual, un lenguaje común. (No quiero negar que en otros contextos la noción de un lenguaje compartido pueda ser muy importante.)

La tesis principal de este trabajo es que una creatura no puede tener pensamientos a menos que sea intérprete del habla de otro. Esta tesis no implica la posibilidad de reducción, conductista o de otro tipo, de pensamiento a habla; ciertamente la tesis no imputa prioridad epistemológica ni conceptual al lenguaje. La idea también difiere de ideas similares en cuanto admite que pueda haber pensamientos para los cuales el hablante no puede encontrar palabras, o para lo cuales no hay palabras.

Alguien que puede interpretar una emisión de la oración castellana "La pistola está cargada" debe tener muchas creencias, y estas creencias deben ser muy parecidas a las creencias que alguien debe tener si considera el pensamiento de que la pistola está cargada. El intérprete, podemos suponer, debe creer que la pistola es un arma, y que es un objeto físico más o menos durable. Probablemente no haya una lista definida de cosas que deba creer alguien que comprenda la oración "La pistola está cargada", pero es necesario que haya innumerables creencias entrelazadas.

Un intérprete conoce las condiciones bajo las cuales emisiones de oraciones son verdaderas, y a menudo sabe que si determinadas oraciones son verdaderas, otras deben serlo. Por ejemplo, un intérprete del castellano sabe que si "La pistola está cargada y la puerta está cerrada" es verdadera, luego "La puerta está cerrada" es verdadera. Las oracio-

nes de un lenguaje tienen una posición en el espacio lógico creado por el patrón de esas relaciones. Resulta obvio que el patrón de relaciones entre oraciones es muy parecido al patrón de relaciones entre pensamientos. Este hecho ha reforzado la visión de que es redundante tomar ambos patrones como básicos. Si los pensamientos son primarios, un lenguaje parece no servir a otro propósito que el de expresar o conducir pensamientos; mientras que si consideramos que el habla es primaria, resulta tentador analizar los pensamientos como disposiciones de habla: como dice Sellars, "... pensar al nivel definidamente humano... es esencialmente actividad verbal".[2] Pero está claro que el paralelo entre la estructura de los pensamientos y la estructura de las oraciones no proporciona argumentos para la primacía de ninguna, y sólo una presunción en favor de su interdependencia.

Hemos estado hablando libremente de pensamientos, creencias, significados e interpretaciones; o más bien, usando libremente oraciones que contienen estas palabras. Pero por supuesto que no está claro qué entidades, o clases de entidades, debe haber para tener una comprensión sistemática de esas oraciones. Sin embargo, el habla aparente de pensamientos y de dichos sí pertenece a un modo familiar de explicación de la conducta humana y debe considerarse una sección organizada del sentido común que puede también ser llamada teoría. Una vía para examinar la relación entre pensamiento y lenguaje pasa por inspeccionar la teoría implícita en este tipo de explicación.

Parte de la teoría trata de la explicación teleológica de la acción. Nos preguntamos por qué un hombre levanta su brazo; una explicación podría ser que quería atraer la atención de un amigo. Esta explicación fracasaría si quien levantó el brazo no creyera que al levantar el brazo atraería la atención de su amigo, por lo que la explicación completa de su acto de levantar el brazo, o al menos una explicación más completa, sería que quería atraer la atención de su amigo *y* creía que al levantar el brazo atraería la atención de su amigo. Una explicación de este tipo familiar tiene algunos rasgos que vale la pena enfatizar. Explica lo que es relativamente aparente —levantar un brazo— apelando a factores que son mucho más problemáticos: deseos y creencias. Pero si fuéramos a pedir evidencia de que la explicación es correcta, esta evidencia consistiría al final en más datos referidos al tipo de hecho que se está aplicando, esto es, más conducta que se explica mediante las creencias y deseos postulados. Hacer referencia a creencias y deseos para explicar la acción es así una forma de ajustar una acción dentro de un patrón de conducta que tiene coherencia gracias a una teoría. Esto no significa, desde luego, que las creencias no sean otra cosa que patrones de conducta, ni que los patrones relevantes puedan ser definidos sin usar los conceptos de creencia y deseo. No obstante, en cierto sentido las atribuciones de creencia y deseo, y por lo tanto las explicaciones teleoló-

[2] W. Sellars, "Conceptual Change", 82.

gicas de creencia y deseo, sobrevienen a una conducta descrita en términos más amplios.

Una característica de la explicación teleológica no compartida por la explicación en general es la forma en que ella apela al concepto de *razón*. La creencia y el deseo que explican una acción deben ser tales que cualquiera que tuviera esa creencia y ese deseo tendría una razón para actuar de ese modo. Lo que es más, las descripciones de la creencia y del deseo que proporcionamos deben exhibir, en la explicación teleológica, la racionalidad de la acción a la luz del contenido de la creencia y de objeto del deseo.

La relevancia de una explicación teleológica descansa, como se señaló, en su capacidad de descubrir un patrón coherente en la conducta de un agente. Aquí la coherencia incluye la idea de racionalidad tanto en el sentido de que la acción a ser explicada debe ser razonable a la luz de los deseos y creencias asignados, como también en el sentido de que los deseos y creencias asignados deben ajustarse entre sí. La presunción metodológica de racionalidad no hace imposible atribuir pensamientos y acciones irracionales a un agente, pero en cambio impone una carga sobre dichas atribuciones. Debilitamos la inteligibilidad de las atribuciones de pensamientos de todo tipo al punto que no logramos descubrir un patrón consistente de creencias y, finalmente, de acciones, puesto que sólo con referencia al encuadre de un patrón así podemos identificar pensamientos. Si vemos a un hombre tirando de ambos extremos de un trozo de cuerda, podemos decidir que está luchando consigo mismo, que quiere mover la cuerda en direcciones incompatibles. Tal explicación requeriría un respaldo elaborado. No surge ningún problema si la explicación es que él quiere cortar la cuerda.

Desde el punto de vista de alguien que da explicaciones telológicas de las acciones de otro, está claro que no tiene sentido asignar prioridad sea a los deseos o a las creencias. Ambos son esenciales para la explicación de la conducta, y ninguno está más directamente abierto a la observación que el otro. Esto crea un problema, pues significa que la conducta, que es la principal base evidencial para las atribuciones de creencia y deseo, se considera el resultado de dos fuerzas menos abiertas a la observación pública. Así, donde una constelación de creencias y deseos racionalizará una acción, siempre es posible hallar una constelación muy diferente que funcionaría igualmente bien. Incluso un muestrario generoso de acciones amenaza con dejar abierto un número inaceptablemente grande de explicaciones alternativas.

Por fortuna disponemos de una teoría más refinada, una todavía firmemente basada en el sentido común: la teoría de la preferencia, o de la toma de decisiones en condiciones de incertidumbre. Frank Ramsey precisó esta teoría por primera vez, aunque él la veía como una forma de proveer fundamento al concepto de probabilidad más que como una porción de psicología filosófica.[3] La teoría de Ramsey actúa cuantifican-

[3] F. P. Ramsey, "Truth and Probability".

do la fuerza de la preferencia y el grado de creencia de manera tal que permite comprender la idea natural de que al elegir un curso de acción consideramos no sólo cuán deseables resultan diversos resultados, sino también cuán aptos para producir esos resultados son los cursos de acción disponibles. La teoría no supone que podemos juzgar los grados de creencia o hacer comparaciones numéricas de valor directamente. Antes bien, ella postula un razonable patrón de preferencias entre cursos de acción, y muestra cómo construir un sistema de creencias y deseos cuantificados para explicar las elecciones. Dadas las condiciones idealizadas que postula la teoría, el método de Ramsey hace posible identificar únicamente las creencias y deseos relevantes. En vez de hablar de postulación, podríamos presentar la cosa así: dentro del límite en que podamos ver que las acciones de un agente siguen los lineamientos de un patrón consistente (racional) de un tipo determinado, podemos explicar dichas acciones en términos de un sistema de creencias y deseos cuantificados.

Volveremos ahora a la teoría de la decisión; es tiempo de inclinarnos hacia la cuestión de cómo se interpreta el habla. La meta inmediata de una teoría de la interpretación es dar el significado de una emisión arbitraria de un miembro de una comunidad de lenguaje. He discutido que para la interpretación es fundamental una teoría de la verdad que satisfaga la Convención T de Tarski (modificada en ciertos aspectos para que sea aplicable a un lenguaje natural). Puede considerarse que una teoría así da una interpretación para cada oración que un hablante podría emitir. Para pertenecer a una comunidad de habla —para ser un intérprete del habla de los demás— uno necesita, en efecto, conocer una teoría de este tipo, y saber que es una teoría del tipo correcto.[4]

Una teoría de la interpretación, al igual que una teoría de la acción, nos permite redescribir ciertos sucesos de una manera reveladora. Así como una teoría de la acción puede responder a la pregunta acerca de qué está haciendo un agente cuando ha levantado su brazo redescribiendo el acto como un intento de llamar la atención de su amigo, igualmente un método de interpretación puede llevar a redescribir la emisión de ciertos sonidos como un acto de decir que la nieve es blanca. Llegados a este punto, empero, la analogía se interrumpe. Pues la teoría de la decisión también puede explicar acciones, mientras que no resulta nada claro de qué manera una teoría de la interpretación puede explicar por qué un hablante emite las palabras "La nieve es blanca". Pero esto, después de todo, era de esperar, pues emitir palabras es una acción, y por lo tanto debe recurrir a las creencias y los deseos para su explicación teleológica. La interpretación no es irrelevante para la explicación teleológica del habla, dado que para explicar por qué alguien dijo que necesitamos conocer, entre otras cosas, su propia interpretación de lo que dijo, esto es, lo que él cree que significan sus palabras en las circunstancias bajo las cuales habla. Naturalmente esto involucrará algunas de sus creencias acerca de cómo los demás interpretarán sus palabras.

[4] Véanse los Ensayos 9 y 10.

La interrelación de la teoría de la acción con la interpretación emergerá de otra manera si preguntamos cómo se pone a prueba un método de interpretación. Al final, la respuesta debe ser que ayuda a poner orden en nuestra comprensión de la conducta. Pero en una etapa intermedia, podemos ver que la actitud de *considerar verdadero* o *aceptar como verdadero*, en referencia a las oraciones, debe desempeñar un papel central para darle forma a la teoría. Por una parte, la mayoría de los usos del lenguaje nos dice directamente, o arroja luz a la cuestión, de si un hablante considera que una oración es verdadera. Si el propósito de un hablante es dar información, o hacer una aserción honesta, entonces normalmente el hablante cree que está emitiendo una oración verdadera bajo esas circunstancias. Si emite una orden, usualmente podemos entender que esto muestra que él considera que determinada oración (cercanamente relacionada a la oración emitida) es falsa; lo mismo ocurre para muchos casos de engaño. Cuando se formula una pregunta, ella generalmente indica que el preguntador no sabe si determinada oración es verdadera; y así por el estilo. Para inferir de esa evidencia que un hablante considera verdadera una oración necesitamos saber mucho acerca de sus deseos y creencias, pero no tenemos que saber lo que significan sus palabras.

Por la otra parte, el conocimiento de las circunstancias bajo las cuales alguien considera que ciertas oraciones son verdaderas es esencial para la interpretación. Vimos en el caso de los pensamientos que si bien la mayor parte de los pensamientos no son creencias, es el patrón de creencia lo que nos permite identificar cualquier pensamiento; análogamente, en el caso del lenguaje, si bien la mayor parte de las emisiones no conciernen a la verdad, es el patrón de las oraciones consideradas verdaderas lo que da a las oraciones su significado.

La actitud de considerar verdadera una oración (bajo circunstancia especificadas) relaciona creencia e interpretación de una manera fundamental. Podemos saber que un hablante considera que una oración es verdadera sin saber qué significa con ella o qué creencia ella expresa para él. Pero si sabemos que él considera verdadera la oración *y* sabemos cómo interpretarla, entonces podemos hacer una correcta atribución de creencia. Simétricamente, si sabemos qué creencia expresa una oración considerada verdadera, sabemos cómo interpretarla. El problema metodológico de la interpretación consiste en ver, dadas las oraciones que un hombre acepta como verdaderas bajo circunstancias dadas, cómo determinar cuáles son sus creencias y qué significan sus palabras. La situación es nuevamente similar a la situación de la teoría de la decisión donde, dadas las preferencias de un hombre entre cursos de acción alternativos, podemos discernir tanto sus creencias como sus deseos. Por supuesto, no debería pensarse que una teoría de la interpretación se sostendrá por sí sola pues, como hemos visto, no hay posibilidad de determinar cuándo una oración es considerada verdadera si no somos capaces de atribuir deseos y de describir a las acciones como portadoras de intenciones complejas. Esta observación no priva de interés a la teoría

de la interpretación, pero le asigna un lugar dentro de una más amplia teoría de la acción y del pensamiento.[5]

Todavía no está claro si se quiere la interpretación para una teoría de la acción, cuestión que nos hemos puesto a responder. Lo que es seguro es que todas las formas habituales de poner a prueba teorías de la decisión o de la preferencia bajo condiciones de incertidumbre confían en el uso del lenguaje. Es relativamente simple eliminar la necesidad de respuestas verbales por parte del sujeto: se puede entender que ha expresado una preferencia ejecutando una acción, moviéndose directamente para alcanzar su fin, en vez de decir lo que quiere. Pero esto no puede resolver la cuestión de lo que elegido. Un hombre que toma una manzana y no una pera cuando se le ofrecen ambas puede estar expresando una preferencia por lo que está a su izquierda y no a su derecha, por lo que es rojo y no amarillo, por lo que ha visto primero, o juzgado más costoso. Si se repiten los tests pueden hacerse algunas lecturas de sus acciones más plausibles que otras, pero restará el problema de determinar cuándo él juzga que dos objetos de elección son idénticos. Los tests que involucran sucesos inciertos —elecciones entre apuestas— son aun más dificiles de presentar sin emplear palabras. El psicólogo, escéptico respecto de su capacidad de entender cómo está interpretando sus instrucciones un sujeto, debe agregar una teoría de la interpretación verbal a la teoría que se pone a prueba. Si pensamos que todas las elecciones revelan una preferencia por la verdad de una oración y no de otra, la teoría total resultante debería proporcionar una interpretación de oraciones, y al mismo tiempo asignar creencias y deseos, entendiéndose que estos últimos relacionan al agente con oraciones o emisiones. Esta teoría compuesta explicaría toda conducta, verbal y de otro tipo.

Todo esto sugiere poderosamente que la atribución de deseos y creencias (y de otros pensamientos) debe avanzar mano a mano con la interpretación del habla, y que ni la teoría de la decisión ni la de la interpretación pueden desarrollarse exitosamente una sin la otra. Pero resta decir, más detallada y convincentemente, por qué la atribución de pensamiento depende de la interpretación del habla. La razón general, no muy informativa, dice que sin el habla no podemos efectuar las distinciones finas entre pensamientos que son esenciales para las explicaciones que a veces podemos proveer con confianza. Nuestro estilo de atribuir actitudes asegura que pueda usarse todo el poder expresivo del lenguaje para hacer tales distinciones. Uno puede creer que Scott no es el autor de *Waverley* pero no dudar de que Scott es Scott; uno puede querer ser el descubridor de una creatura con corazón sin querer ser el descubridor de una creatura con riñones. Uno puede pretender morder la manzana que tiene en la mano sin pretender morder la única manzana que tiene dentro un gusano; y así sucesivamente. La intensionalidad

[5] El entrelazamiento entre la teoría de la decisión y la interpretación radical se explora también en el Ensayo 10, en el Ensayo 12 de *Essays on Trought and Action*, y en "Toward a Unified Theory of Meaning and Action".

que compredemos tan bien en la atribución de pensamientos es muy difícil de comprender cuando el habla no está presente. El perro, decimos, sabe que su dueño está en casa. ¿Pero sabe que el señor Smith (que es su dueño), o que el presidente del banco (que es ese mismo dueño), está en casa? No tenemos una idea real de la forma de resolver, o comprender, estas preguntas. Es mucho más difícil decir, cuando el habla no está presente, cómo distinguir pensamientos universales de conjunciones de pensamientos, o cómo atribuir pensamientos condicionales, o pensamientos con, por ejemplo, cuantificación mixta ("El anhela que todos sean amados por alguien").

Estas consideraciones probablemente serán menos persuasivas para los amantes de los perros que para los demás, pero en cualquier caso no constituyen un argumento. A lo que sumo lo que hemos mostrado, o pretendido decir, es que a menos que haya conducta que pueda interpretarse como habla, la evidencia no resultará adecuada para justificar las distinciones finas que estamos acostumbrados a hacer en la atribución de pensamientos. Si persistimos en atribuir deseos, creencia u otras actitudes bajo estas condiciones, nuestras atribuciones y consecuentes explicaciones de acciones se verán seriamente indeterminadas por cuanto muchos sistemas alternativos de atribución, muchas explicaciones aternativas, estarán igualmente justificados por los datos disponibles. Quizás esto sea todo cuanto podamos decir contra la atribución de pensamientos a los animales; pero yo no lo creo.

Antes de proseguir quiero considerar una posible objeción a la línea general que he estado impulsando. Supóngase que concedemos, dice el impugnador que la conducta muy compleja no observada en los infantes y en los elefantes resulta necesaria si es que queremos encontrar aplicación a todo el aparato disponible para la atribución de pensamientos. Puede decirse aún que el esquema de cómo funciona la interpretación no muestra que esta complejidad deba entenderse como relacionada con el lenguaje. La razón está en que dicho esquema hace que sea demasiado lo que depende de la actitud especial de ser pensado como verdadero. La evidencia más directa de la existencia de esta actitud es la aserción honesta. Pero entonces parecería que podríamos tratar como habla a la conducta de creaturas que nunca hicieron nada con el lenguaje excepto hacer aserciones honestas. Algunos filósofos efectivamente sueñan con semejantes tribus monocordes; pero ¿sería correcto decir que tienen un lenguaje? Se ha perdido de vista lo que puede llamarse *la autonomía del significado*. Una vez que una oración se ha comprendido, una emisión de ella puede usarse para servir a casi cualquier propósito extralingüístico. Un instrumento al que pudiera darse solamente un uso único carecería de autonomía de significado; esto equivale a decir que no debe ser tenido en cuenta como un lenguaje. Por lo tanto, la complejidad de conducta que hace falta para dar la máxima extensión a las atribuciones de pensamiento no necesita, después de todo, tener exactamente la misma complejidad que admite, o requiere, la interpretación como un lenguaje.

Coincido con el hipotético impugnador en que la autonomía de sig-

nificado es esencial para el lenguaje; por cierto, es esto mayormente lo que explica por qué el significado lingüístico no puede ser definido o analizado en base a intenciones y creencias extralingüísticas. Pero el impugnador no logra distinguir entre un lenguaje que *podría* usarse para un solo propósito y uno que *se* usa para un solo propósito. Un instrumento que podría usarse para un solo propósito no sería lenguaje. Pero la aserción honesta, sola, podría producir una teoría de la interpretación, e igualmente un lenguaje al que, si bien capaz de más, nunca se le diera otros usos. (En la práctica, este suceso es impensable. Alguien que conoce las condiciones bajo la cuales sus oraciones son socialmente verdaderas no puede dejar de echar mano a, y aprovecharse de, las posibilidades de la aserción deshonesta: sea para bromear, contar historias, estimular, exagerar, insultar, y demás.)

Un método de interpretación nos dice que para hablantes del castellano una emisión de "Está lloviendo" por un hablante x es un tiempo t es verdadera si y sólo si está lloviendo (cerca de x) en t. Estar provisto de esta información, y conocer que otros la conocen, equivale a conocer lo que una emisión significa, independientemente del conocimiento de los propósitos que le dieron impulso. La autonomía de significado también ayuda a explicar cómo es posible, mediante el uso del lenguaje, atribuir pensamientos. Supongamos que alguien emite asertivamente la oración "La nieve es blanca". Conociendo las condiciones bajo la cuales dicha emisión es verdadera yo puedo agregar, si lo deseo, "Yo también creo *that*", atribuyéndome de esa manera una creencia. En este caso ambos podemos haber afirmado que la nieve es blanca, pero la igualdad de fuerza no es necesaria para la autoatribución. El otro puede decir con una mueca, expresando descrédito, "La nieve es blanca"; y yo puedo nuevamente atribuirme una creencia diciendo, "Pero *yo* creo *that*". Esto también puede funcionar de otra manera: si yo puedo aprovechar una emisión de otro para atribuirme una creencia, puedo usar una emisión mía para atribuir una creencia a otro. Primero emito una oración, tal vez "La nieve es blanca", y después agrego "El cree *that*". La primera emisión puede o no ser una aserción; en todo caso, no atribuye una creencia a nadie (aunque si es una aserción, luego yo me *represento* a mí mismo creyendo que la nieve es blanca). Pero si mi observación "El cree *that*" es una aserción, he atribuido una creencia a alguien. Finalmente, no hay ningún impedimento para que yo me atribuya una creencia diciendo primero, "La nieve es blanca" y agregando luego, "Yo creo *that*".

En todos estos ejemplos, tomo la palabra "*that*" para referir demostrativamente una emisión, sea una emisión del hablante del "*that*" o de otro hablante. El "*that*" no puede referir una oración: tanto porque, como Church ha señalado en casos similares, la referencia entonces tendría que relativizarse a un lenguaje pues una oración puede tener diferentes significados en diferentes lenguajes,[6] porque, de modo más obvio,

[6] A. Church, "On Carnap's Analysis of Statements of Assertion and Belief".

la misma oración puede tener diferentes valores de verdad en el mismo lenguaje.

Lo que la referencia demostrativa a las emisiones hace en el tipo de caso que acabamos de considerar puede hacerlo igualmente cuando la estructura de superficie se ha alterado hacia algo como "Yo creo *that* la nieve es blanca" o "El cree *that* la nieve es blanca". También pienso que en estas instancia debemos ver al *"that"* como un demostrativo, refiriendo ahora una emisión a punto de producirse. Así, la forma lógica de la atribuciones normales de actitud es la de dos emisiones unidas paratácticamente. No hay conectivo, si bien la primera emisión contiene una referencia a la segunda. (Observaciones similares valen, desde luego, para inscripciones de oraciones.)

He discutido este análisis de atribuciones verbales de actitud en otra parte, y no hay necesidad de repetir aquí los argumentos y explicaciones.[7] Es un análisis con sus propias dificultades, especialmente cuando se trata de analizar la cuantificación dentro de la oración contenida, pero pienso que podemos superar estas dificultades preservando los rasgos atractivos de la idea. Aquí quiero insistir en un punto que conecta el análisis paratáctico de las atribuciones de actitud con nuestro tema presente. El análisis propuesto relaciona directamente la característica autónoma de significado con nuestra capacidad de describir y atribuir pensamientos, pues la emisión de una oración puede servir para la descripción de las actitudes de otros solamente por causa de que la interpretación de una oración es independiente de su uso. Si mi análisis es correcto, podemos olvidarnos de la improbable (pero común) opinión de que una oración encerrada dentro de una *cláusula-"that"* necesita una interpretación completamente diferente de aquella que la satisface en otros contextos. Dado que en contextos ordinarios las oraciones no son nombres ni descripciones, podemos rechazar en particular la presunción según la cual las actitudes tienen objetos tales como proposiciones que se considerarían nombrados o descritos por las *cláusulas-"that"*. Tendría que evitarse la tentación de llamar objeto de la actitud atribuida a la emisión a la cual se hace referencia de acuerdo con el análisis paratáctico.

Aquí, una solución fácil a nuestro problema acerca de la relación entre pensamientos y habla se sugiere a sí misma. Una forma de ver el análisis paratáctico, forma propuesta por Quine en *Word and Object*, es la siguiente: cuando un hablante atribuye una actitud a una persona, lo que hace es imitar o remedar una acto de habla real o posible de esa persona.[8] El discurso indirecto es el mejor ejemplo, y otro buen ejemplo lo constituye la aserción. Supóngase que yo digo: "Herodoto afirmó que el Nilo crece en las Montañas de la Luna". Mi segunda emisión —mi emisión apenas hecha de "El Nilo crece en las Montañas de la Luna"— debe conllevar, si mi atribución a Herodoto es correcta, una determina-

7 Véase el Ensayo 7.
8 W. V. Quine, *Word and Object*, 219.

da relación con una emisión de Herodoto: debe ser, en algún sentido apropiado, una traducción de ella. Dado que (suponiendo aun que la atribución es correcta) Herodoto y yo somos *igual-dicentes*, mi misión imitaba la suya. No respecto de la fuerza, desde luego, puesto que yo no afirmé nada acerca del Nilo. La igualdad se da respecto del contenido de nuestras emisiones. Si hablamos de otras actitudes, la situación es más complicada, pues característicamente no hay emisiones para imitar. Si yo afirmo "Jones cree que la nieve es blanca", mi emisión de "La nieve es blanca" puede no tener ninguna emisión real de Jones para imitar. Sin embargo, podemos entender aun que lo que yo afirmo es que Jones estaría diciendo honestamente lo que piensa si emitiera una oración que tradujera la mía. Dadas algunas delicadas suposiciones acerca de las condiciones bajo las cuales tal condicional subjuntiva es verdadero, podríamos concluir que sólo alguien con un lenguaje podría tener un pensamiento, puesto que tener un pensamiento sería tener una disposición para emitir determinadas oraciones con la fuerza apropiada bajo circunstancias dadas.

Podríamos tomar este camino, pero desafortunadamente no parece haber una razón clara de por qué tendríamos que hacerlo. Nos hemos puesto en marcha para hallar un argumento que muestre que solamente las creaturas con habla tienen pensamientos. Lo que se acaba de delinear no es un argumento sino una propuesta, y no necesitamos aceptar una propuesta. El análisis paratáctico de la forma lógica de las atribuciones de actitud puede pasárselas sin la teoría imitativa de la emisión. Cuando yo digo "Jones cree que la nieve es blanca" describo directamente la condición mental de Jones: ciertamente se trata de la condición mental en la que se encuentra alguien que podría afirmar honestamente "La nieve es blanca" si hablara en castellano, pero ésa podría ser una condición en la que también podría encontrarse una creatura sin lenguaje.

Para poder llegar a la última cuestión importante, debo volver a un aspecto de la interpretacióin que hasta ahora hemos dejado de lado. Señalé que la actitud de considerar verdadero, dirigida a oraciones bajo circunstancias especificadas, es la base de la interpretación, pero no dije de qué manera puede ella cumplir esta función. La dificultad, se recordará, consiste en que una oración se considera verdadera por causa de dos factores: lo que el evaluador entiende que significa la oración, y lo que él cree. Para que sea posible ordenar las cosas, necesitamos un método para mantener fijo un factor mientras se estudia el otro.

La condición de miembro de una comunidad de lenguaje depende de la capacidad de interpretar las emisiones de los miembros del grupo, y disponemos de un método para lograrlo si uno tiene, y sabe que tiene, una teoría que provee condiciones de verdad,más o menos al estilo de Tarski, para todas las oraciones (relativizadas, como siempre, a un tiempo y un hablante). La teoría es correcta siempre y cuando implique, por medios finitamente enunciados, teoremas de la forma familiar: " 'Está lloviendo' es verdadero para un hablante x en un tiempo t si y sólo si está lloviendo (cerca de x) en t". La base evidencial para una teoría

así interesa oraciones consideradas verdaderas, hechos como el siguiente: " 'Está lloviendo' fue considerado verdadero por Smith a las 8 de la mañana del 26 de agosto y efectivamente llovía cerca de Smith en ese momento". Sería posible generar una teoría correcta simplemente registrando como verdaderas las oraciones consideradas verdaderas, siempre que (1) hubiera una teoría que satisficiera las restricciones formales y fuera consistente de esta manera con la evidencia, y (2) todos los hablantes consideran que una oración es verdadera precisamente cuando esa oración fuera verdadera; con la condición de que todas las creencias, al menos hasta donde pudieran expresarse, fueran correctas.

Pero naturalmente no puede suponerse que los hablantes jamás tienen creencia falsas. El terror da su peculiaridad a la creencia. Sin embargo podemos entender que *la mayoría* de las creencias es correcta. La razón de esto es que una creencia se identifica por su posición en un patrón de creencias; este patrón es el que determina la materia de la creencia, de qué trata la creencia. Antes de que un objeto en el mundo, o un aspecto del mundo, pueda formar parte de la materia de que trata una creencia (verdadera o falsa) debe haber innumerables creencias verdaderas acerca de esa materia. Las creencia falsas tienden a socavar la identificación de la materia; socavan, por lo tanto, la validez de una descripción de la creencia que la vincula con esa materia. Y así, a su turno, las creencias falsas socavan la afirmación de que una creencia conexa es falsa. Para tomar un ejemplo, ¿qué tan seguros estamos de que los antiguos —algunos antiguos— creían que la Tierra era plana? *¿Esta* Tierra" Bueno, esta Tierra nuestra es parte del sistema solar, un sistema identificado parcialmente por el hecho de ser un racimo de cuerpos sólidos, grandes y fríos, que gira en torno de una estrella muy grande y caliente. Si alguien *no* cree *nada* de esto acerca de la Tierra, ¿es seguro que está pensando en la Tierra? No se requiere una respuesta. La cuestión es, si este tipo de consideración de creencias relacionadas puede sacudir nuestra seguridad de que los antiguos creían que la Tierra era plana. No se trata de que cada creencia falsa destruya necesariamente nuestra capacidad de identificar otras creencias, sino de que la inteligibilidad de esas identificaciones debe depender de un entorno de creencias verdaderas mayormente no mencionadas y no cuestionadas. Para decirlo de otra forma: cuanto para más cosas está en lo cierto el creyente, más definidos son sus errores. Un exceso de error simplemente empaña el foco.

Lo que hace la interpretación posible, entonces, es el hecho de que podemos descartar *a priori* la posibilidad de un error generalizado. No puede ser correcta una teoría de la interpretación que haga a un hombre asentir ante una gran cantidad de oraciones falsas: generalmente debe darse el caso de que una oración es verdadera cuando un hablante sostiene que lo es. Tanto es así, que se computa a favor de un método de interpretación el que éste cuente a una oración como verdadera precisamente cuando los hablantes la consideran verdadera. Pero, desde luego, el hablante puede estar equivocado; y también puede estarlo el intérprete. Entonces, lo que finalmente debe contarse en favor de un método de

interpretación es que logre que el intérprete llegue a un acuerdo general con el hablante: según el método, el hablante considera verdadera a una oración bajo condiciones especificadas, y estas condiciones se presentan, en la opinión del intérprete, precisamente cuando el hablante considera que la oración es verdadera.

Ninguna teoría simple puede lograr que el hablante y el intérprete lleguen a un acuerdo perfecto, y por lo tanto una teoría practicable de vez en cuando debe suponer un error por parte de uno o de otro. El precepto metodológico básico es, entonces, que una buena teoría de la interpretación maximice el acuerdo. O, dado que el número de oraciones es infinito, y dadas otras consideraciones por venir, una palabra mejor sería *optimizar*.

Algunos desacuerdos son más destructivos de la comprensión que otros, y como es natural una teoría elaborada debe tener esto en cuenta. El desacuerdo respecto de cuestiones teóricas puede (en algunos casos) ser más tolerable que el desacuerdo respecto de lo que es más evidente; un desacuerdo respecto de cómo se ven o parecen las cosas es menos tolerable que un desacuerdo respecto de cómo son; un desacuerdo acerca de la verdad de atribuciones de determinadas actitudes a un hablante por parte de ese mismo hablante puede no ser tolerable en modo alguno, o serlo apenas. Resulta imposible ejemplificar las consideraciones que son relevantes, pues todo aquello que conocemos o creemos acerca de la forma en que la evidencia apoya a la creencia puede volcarse a la tarea de decidir en qué punto puede la teoría admitir el error, y qué errores son menos destructivos para la comprensión. A este respecto, la metodología de interpretación no es otra cosa que epistemología en el espejo del significado.

El intérprete que presuma que su método puede ponerse en funciones para una comunidad de lenguaje se esforzará por dar con una teoría que optimice un acuerdo en toda la comunidad. Puesto que una comunicación fácil tiene un valor de supervivencia, ese intérprete puede esperar que el uso dentro de una comunidad conduzca a teorías de la interpretación simples y comunes.

Si esta consideración de la interpretación radical es correcta, al menos en líneas generales, entonces deberíamos consignar que los conceptos de verdad objetiva, y de error, necesariamente emergen en el contexto de interpretación. La distinción entre una oración que se considera verdadera y una que es efectivamente verdadera es esencial para la existencia de un sistema interpersonal de comunicación y, cuando en casos individuales se presenta una diferencia, debe contarse como un error. Dado que la actitud de considerar verdadero es la misma, sea la oración verdadera o no, aquélla corresponde directamente a la creencia. El concepto de creencia está así preparado para llenar el vacío entre la verdad objetiva y lo considerado verdadero, y llegamos a entenderlo precisamente en esta conexión.

Poseemos la idea de creencia solamente a partir del rol de la creencia en la interpretación del lenguaje, pues como actitud privada ella no es inteligible excepto como ajuste a la norma pública provista por el len-

guaje. De allí se desprende que una creatura debe ser miembro de una comunidad de habla si es que va a poseer el concepto de creencia. Y dada la dependencia de otras actitudes respecto de la creencia, podemos decir, generalizando, que sólo una creatura que pueda interpretar el habla puede poseer el concepto de un pensamiento.

¿Puede una creatura tener una creencia si no tiene el concepto de creencia? A mí me parece que no, y por esta razón. Una persona no puede tener una creencia a menos que comprenda la posibilidad de estar equivocada, y esto requiere captar el contraste entre verdad y error: creencia verdadera y creencia falsa. Pero, he argumentado, este contraste sólo puede emerger en el contexto de la interpretación, que por sí mismo nos empuja hacia la idea de una verdad objetiva, pública.

A menudo se piensa equivocadamente que el concepto semántico de verdad es redundante, que no hay diferencia entre afirmar que una oración *s* es verdadera y usar *s* para hacer una aserción. Puede sí ser correcta una teoría de redundancia de la creencia, que creer que *p* pueda no distinguirse de la creencia de que *p* es verdadero. Esta noción de verdad no es la noción semántica: el lenguaje no está directamente presente en la ilustración.

Pero está justo fuera de la ilustración; es parte de su marco. Pues la noción de una creencia verdadera depende de la noción de una emisión verdadera, y ésta a su vez no puede existir sin un lenguaje compartido. Como dice el Ulises, de Shakespeare:

> *"... no man is the lord of anything,*
> *Though in and of him there be much consisting,*
> *Till he communicate his parts to others;*
> *Nor doth he of himself know them for aught*
> *Till he behod them formed in th'applause*
> *Where they're extended."*

(*Troilus and Cressida*, III. iii. 115-20)

12

Respuesta a Foster

Hay mucho con lo que yo concuerdo y, aun más, admiro, en el trabajo del señor Foster. Comparto su predisposición en favor de los lenguajes extensionales de primer orden; me complace acompañarlo en la búsqueda de una teoría explícitamente semántica que considere recursivamente los significados de las oraciones en términos de sus estructuras; y me alegra que él coincida en sostener que una teoría puede juzgarse adecuadamente en base a restricciones holísticas. Especialmente aplaudo a Foster por lo que él deja de lado: tal como Lear obtiene el poder merced a la ausencia de Cordelia, pienso que el tratamiento del lenguaje prospera cuando evita la evocación acrítica de los conceptos de convención, regla lingüística, práctica lingüística o juegos de lenguaje.

Aun en el lado positivo, pienso que Foster está en lo cierto cuando pregunta si una teoría propuesta enuncia explícitamente algo cuyo conocimiento bastaría para interpretar emisiones de hablantes del lenguaje al cual ella se aplica. (Evito la palabra "maestría", y la competencia especial de un hablante, por razones que, creo, no afectarán nuestra discusión.) Tardé en apreciar la importancia de esta manera de formular una finalidad general de las teorías del significado, si bien hay elementos de la idea que aparecen en varios de mis trabajos tempranos.[1] Agradezco a varios amigos de Oxford el que me hayan urgido a tratar de clarificar mis ideas sobre este tema: aquí tengo que mencionar especialmente a Michal Dummett, Gareth Evans, John McDowell y John Foster.

En un trabajo que se leyó por primera vez en Biel, Suiza, en mayo de 1973, critiqué mis propios intentos anteriores de decir exactamente cuál es la relación entre una teoría de la verdad y una teoría del significado, y traté en él de hacerlo mejor.[2] Leí nuevamente este trabajo en Windsor (noviembre de 1973), y se convirtió en la base de abundantes discusiones en un seminario que Michael Dummett y yo dimos en Oxford en Trinity Term, en 1974. La críticas que introduje allí contra mi formulación previa son (creo yo) esencialmente las que elaborara Foster en la segunda parte del trabajo suyo que aquí citamos, y mi intento de búsqueda de algo mejor se encuentra entre las ideas que él ataca en la tercera parte de su trabajo.

[1] Por ejemplo, los Ensayos 1 y 2.
[2] El trabajo mencionado es el Ensayo 9.

Concuerdo en general con Foster en cuanto a que todavía tengo que ofrecer una formulación completamente satisfactoria de qué cosa es, en mi aproximación, lo que basta conocer para estar en capacidad de interpretar las emisiones de un hablante. Por otro lado, tengo esperanzas de no estar tan lejos de ese objetivo como piensa él, y sus argumentos no me han persuadido de que "mi gran proyecto se encuentre en ruinas". Por cierto que aún me parece correcto, en su medida, sostener que alguien está en posición de interpretar las emisiones de los hablantes de un lenguaje *L* si ese alguien posee un determinado cuerpo de conocimiento implicado por una teoría de la verdad para *L* —una teoría que satisfaga restricciones empíricas y formales especificadas— y sabe que este conocimiento está implicado por dicha teoría.

Tarski dice, y es casi suficiente para nuestros propósitos, que una teoría de la verdad para un lenguaje *L* es satisfactoria siempre que implique (por medio de un conjunto finito de axiomas no lógicos y de lógica normal), para cada oración *s* de *L*, un teorema de la forma:

s es verdadera en *L* si y sólo si *p*.

Donde "*s*" se reemplaza por una descripción estandarizada de *s* y "*p*" se reemplaza por una traducción de *s* al lenguaje de la teoría. Si conociéramos esa teoría, y si supiéramos que se trata de tal teoría, luego podríamos producir una traducción de cada oración de *L*, y sabríamos que se trata de una traducción. Conoceríamos más cosas, pues sabríamos en detalle en qué forma los valores de verdad de las oraciones de *L* se debían a sus estructuras, y por qué algunas oraciones implicaban otras, y cómo las palabras desempeñaban sus funciones a fuerza de relacionarse con objetos del mundo.

Dado que Tarski estaba interesado en definir la verdad, y estaba trabajando con lenguajes artificiales donde la estipulación puede reemplazar a la iluminación, podía dar por sentado el concepto de traducción. Pero en la interpretación *radical*, esto es justamente lo que no puede suponerse. Por ello he propuesto algunas restricciones empíricas para la aceptación de una teoría de la verdad que pueda enunciarse sin apelar a conceptos como significado, traducción o sinonimia, aunque no sin una cierta comprensión de la noción de verdad. He tratado de mostrar por la vía del razonamiento que si una teoría satisface las restricciones, luego las oraciones T que emergen de esa teoría tendrán de hecho traducciones de *s* reemplazando a "*p*".

Aceptar este cambio de perspectiva no equivale a abandonar la Convención T, sino a leerla en una nueva forma. Al igual que Tarski, aspiro a una teoría que satisfaga la Convención T, pero donde él supone el concepto de traducción para poder echar la luz sobre el de verdad, yo quiero iluminar el concepto de traducción por medio de la suposición de una comprensión parcial del concepto de verdad.

Ya al escribir "Verdad y Significado" tenía claro que para que las teorías aceptables de la verdad incluyeran solamente aquellas que servirían para la interpretación, debía agregarse a las restricciones forma-

les esas restricciones empíricas que acabamos de mencionar. Mi error no fue, como parece sugerirlo Foster, suponer que *toda* teoría que diera correctamente condiciones de verdad serviría para la interpretación; mi error consistió en pasar por alto el hecho de que alguien podría conocer una teoría lo suficientemente peculiar sin saber que lo es. No fue extraño que yo descuidara esa distinción, pues imaginé que quien conociera la teoría sería alguien que la habría construido a partir de la evidencia, y tal persona no podía pasar por alto que su teoría satisfacía las restricciones.

Foster nota la diferencia entre dos preguntas que podrían surgir en torno a mi propuesta. Una es si las restricciones que he exigido de una teoría aceptable son adecuadas para asegurar que ella satisfaga la Convención T; por ejemplo, para asegurar que en su oración T la rama derecha del bicondicional realmente traduzca la oración cuyo valor de verdad ella está dando. La otra pregunta es si he logrado decir qué es lo que conoce un intérprete competente (o qué le bastaría conocer). A Foster sólo le interesa aquí la segunda pregunta; él desea acordar, para el espacio del argumento, que las restricciones resultan adecuadas para sus propósitos.

Desde ese punto de vista debemos comprender la discusión de Foster respecto de las teorías de la verdad que dan correctamente la extensión del predicado de verdad —teorías cuyas oraciones T son todas verdaderas— pero que no satisfacen la Convención T. De esta manera Foster opina igual que yo (por el momento) al suponer que mis criterios no admitirán una teoría que contenga como oración T la siguiente:

> "*a* es parte de *b*" es verdadero si y sólo si *a* es parte de *b* y la Tierra se mueve.

Para Foster, la cuestión es más bien que, no obstante mi intérprete tiene una teoría que satisface la Convención T, no hay nada en la teoría misma que se lo haga saber.

La misma cuestión aparece cuando Foster dice:

> como se entiende de ordinario, enunciar las condiciones de verdad de una oración es decir que... es necesario y suficiente para su verdad, es demarcar, dentro del rango total de circunstancias posibles, aquel subconjunto con el cual la oración concuerda. Pero no es éste el sentido en el cual una *oración T* enuncia condiciones de verdad. Una oración T no dice que tal y tal tipo estructural *sería* verdadero... para todas las circunstancias en las que *fue* el caso que... pero meramente que, siendo las cosas como son, este tipo estructural *es* verdadero si y sólo si...

Una teoría que pase los tests empíricos es una que efectivamente puede proyectarse a casos inobservados y contrafácticos, y esto resulta aparente para quienquiera que conozca cuál es la evidencia y cómo se la usa para apoyar la teoría. El problema es que la teoría no enuncia que tiene el carácter que tiene.

Tenemos un paralelo preciso si preguntamos qué datos debe cono-

cer alguien para ser un físico. Una rápida respuesta podría ser: las leyes de la física. Pero Foster diría, y yo estoy de acuerdo, que esto no basta. El físico debe conocer también (y aquí hablo por mí mismo) que esas leyes *son* leyes; digamos, que se ven conformadas por sus instancias, y que soportan afirmaciones contrafácticas y subjuntivas. Para hacerlo más claro, imagine que se le dice a un niño, futuro científico, que la masa de un cuerpo no influye en el tiempo que le lleva caer una distancia dada en el vacío. Luego se le pregunta: "Suponte que Galileo hubiera arrojado una pluma y una bala de cañón desde la cima del Empire State Building, y que la Tierra no tuviera atmósfera. ¿Cuál habría llegado primero al suelo, la pluma o la bala de cañón? " El lúcido niño responde: "No tengo idea. Usted me contó lo que *efectivamente* sucede, siendo las cosas como son; no dijo qué pasaría si las cosas fueran de otra manera".

Foster ofrece, como una tesis que, piensa, yo puedo haber "tratado de impulsar" sin lograr hacerlo correctamente, lo siguiente: "lo que necesitamos conocer, para el dominio de *L*, es tanto los hechos que (una teoría T) enuncia, como que esos hechos tal como los conocemos, son T-teóricos". Luego lo dice en términos simples: lo que alguien necesita conocer es que alguna teoría T para *L* enuncia que... (y aquí los puntos suspensivos se reemplazan por una teoría T). Acepto con agrado esta versión, puesto que es equivalente a la mía. (Hasta donde yo sé, nunca sostuve el punto de vista que él me atribuye y que deja desconcertados el conocimiento de lo que enuncia una teoría de la verdad y el conocimiento de que la teoría es T-teórica.)

Consideremos ahora el punto de vista que Foster piensa que yo debería sostener, y que yo sé que sostengo. No puede decirse que, según este punto de vista, el conocimiento de un lenguaje se reduzca a saber cómo traducirlo a otro. El intérprete, por cierto, sabe que su conocimiento consiste en lo que está enunciado por una teoría T, una teoría T que es traduccional (que satisface la Convención T). Pero no hay razón para suponer que el intérprete puede expresar su conocimiento en alguna forma lingüística específica, y mucho menos en un lenguaje particular.

Quizá deberíamos insistir en que una teoría es una oración o un conjunto de oraciones de algún lenguaje. Pero para conocer una teoría no es ni necesario ni suficiente conocer que estas oraciones son verdaderas. No es suficiente dado que esto podría conocerlo alguien que no tuviera idea de lo que las oraciones significan, y no es necesario pues alcanza con conocer las verdades que expresan las oraciones de la teoría, y esto no requiere el conocimiento del lenguaje de la teoría.

Alguien que puede interpretar el castellano conoce, por ejemplo, que una emisión de la oración "La nieve es blanca" es verdadera si y sólo si la nieve es blanca; sabe además que este hecho lo implica una teoría traduccional: que no es un hecho accidental acerca de esa oración castellana, sino un hecho que *interpreta* la oración. Una vez que el propósito de expresar las cosas así ha quedado claro, no veo ningún inconveniente en reescribir lo que el intérprete conoce en un estilo más familiar: él sabe que "La nieve es blanca" en castellano *significa que* la nieve es blanca.

Está claro, entonces, que mi punto de vista no pretende que la capacidad de interpretar un lenguaje dependa de estar en condiciones de traducir ese lenguaje a una lengua familiar. Tal vez valga la pena reforzar este punto poniendo en su lugar una cuestión que hasta ahora hemos omitido. En los lenguajes naturales los elementos deícticos, como los demostrativos y los tiempos verbales, significan que las condiciones de verdad para muchas oraciones deben hacerse relativas a las circunstancias de su emisión. Cuando esto está hecho, el lado derecho del bicondicional de una oración T nunca traduce la oración para la cual está dando las condiciones de verdad. En general, una teoría adecuada de la verdad no usa recursos deícticos, y por lo tanto puede no contener traducciones de un número y variedad de oraciones muy grande. Con respecto a estas oraciones, ni siquiera existe la ilusión de que la interpretación dependa de la capacidad de traducir. (La forma "significa que" tampoco ayuda aquí.)

Foster piensa que mi gran plan está en ruinas a raíz de que en el intento de utilizar la afirmación de teoricidad T para resguardar la interpretación debo usar una noción intensional como el "enuncia" en "El intérprete sabe que alguna teoría T enuncia *that*...". Pero aquí me impone subrepticiamente una meta que nunca perseguí. Mi camino para tratar de ofrecer una consideración del lenguaje y del significado da un uso fundamental a conceptos tales como creencia e intención, y yo no creo que sea posible reducir estas nociones a algo más científico o conductista. He tratado de dar una consideración del significado (interpretación) que no hiciera un uso esencial de conceptos *lingüísticos* no explicados. (Incluso esto es algo más fuerte de lo que creo posible.) Ninguno de mis planes se arruinará si para decir qué conoce un intérprete es necesario usar una de las llamadas nociones intensionales: una que se asocie a la creencia, a la intención y demás.

Por supuesto que mi proyecto requiere efectivamente que todas las oraciones de lenguajes naturales puedan ser manipuladas por una teoría T, y es por ello que si las formas idiomáticas intensionales resisten tal tratamiento, mi plan se va a pique. Parece ser el caso —si bien la cuestión no es del todo simple o clara— que una teoría de la verdad que satisfaga algo como la Convención T no puede admitir una semántica intensional, y esto me ha urgido a tratar de mostrar de qué manera una semántica extensional puede manipular las peculiaridades de las oraciones de creencia, del discurso indirecto y de otras oraciones así. Foster piensa que mi análisis no dará resultado, pero no resulta fácil ver en qué forma esto es relevante para nuestro debate. Si su criterio es que *ninguna* teoría T puede ofrecer una semántica satisfactoria para oraciones que atribuyen actitudes, entonces toda la discusión acerca de cómo describir exactamente la competencia de un hablante es simplemente irrelevante. Pero si algún análisis es posible, el mío u otro, luego lo que funcione para el discurso indirecto y para las oraciones acerca de la creencia y la intención presumiblemente funcionará también para la relación de "enuncia" que preocupa a Foster.

Ciertamente Foster está en lo cierto cuando dice que la expresión

"una teoría T enuncia *that*" es lo que usualmente se llamaría un operador oracional no veritativo funcional, pues cuando se lo sigue con oraciones materialmente equivalentes puede producir resultados con valores de verdad divergentes. Esto nos acarrea dos problemas (a los cuales quizá Foster no distingue suficientemente entre sí). El primero es si el análisis paratáctico del discurso indirecto puede ser aplicado correctamente al presente caso; el otro es si mi consideración de la interpretación radical se ve amenazada en caso de que la noción relevante de enunciación (sea cual fuere su semántica) oculta un concepto lingüístico inanalizado. El primer problema es, como acabo de sugerir, sólo marginalmente afín a nuestra discusión; el segundo es obviamente esencial. Me gustaría, empero, decir algo sobre ambos menos tópicos, pues a menos que se demuestre que un análisis así es defectuoso yo propongo una semántica paratáctica para "enuncia *that*"

La aproximación semántica paratáctica al discurso indirecto nos indica entender que una emisión de "Galileo dijo *that* la Tierra se mueve" consiste en la emisión de dos oraciones, "Galileo dijo *that*" y "La Tierra se mueve". El "*that*" refiere a la segunda emisión, y la primera emisión es verdadera si y sólo si una emisión de Galileo fue igual en contenido a ("traduce") la emisión a la cual el "*that*" refiere. (Foster dice erróneamente que mi análisis de "Galileo dijo *'that'* " es "Alguna emisión de Galileo y mi última emisión hacen que Galileo y yo seamos igual-dicentes". Esto no es un análisis, sino una reescritura concebida para dar al lector una percepción de la semántica; un recurso expositivo y heurístico.)

Foster intenta probar que mi análisis semántico es erróneo mostrando que fracasa en un test de traducción. Este test exige que la traducción de una emisión (cuando se la analiza) debe establecer el mismo hecho o proposición (estoy empleando las palabras de Foster) que la emisión original. Luego señala que una traducción de "Galileo dijo *that*" al francés, que preservara la referencia de "*that*" (en mi análisis) no lograría comunicar a una audiencia francesa nada acerca del contenido de la afirmación de Galileo. Foster entiende que esto evidencia que, en mi análisis, una emisión ordinaria de una oración como "Galileo dijo *that* la Tierra se mueve" no logra enunciar lo que dijo Galileo; y por lo tanto un análisis paralelo del "enuncia *that*" que necesita mi teoría de la interpretación sufriría el mismo fracaso.

¿Pero qué es esta relación entre emisiones, de establecer el mismo hecho o proposición, que Foster tiene en mente? Todo lo que nos dice es que la referencia debe preservarse. Mas esto seguramente no es suficiente, pues si dos emisiones cualesquiera enuncian el mismo hecho cuando la referencia se preserva, resulta muy difícil bloquear una prueba familiar de que todas las emisiones verdaderas enuncian el mismo hecho. El uso de la palabra "proposición" sugiere que al igual que la referencia debe preservarse el significado. Pero si ambos, referencia y significado, deben preservarse, es fácil ver que muy pocos pares de emisiones pueden enunciar el mismo hecho cuando las emisiones contengan expresiones deícticas. Si dejamos de lado los bilingües, no hay una emi-

sión del francés que pueda enunciar un hecho que enuncio yo al usar "yo", y yo no puedo enunciar dos veces el mismo hecho cuando digo "Yo tengo calor" dos veces. Juzgar erróneo mi análisis en base a estas pautas es simplemente juzgarlo erróneo porque supone que el discurso indirecto involucra un elemento deíctico. A falta de otros argumentos, la conclusión genera dudas respecto a las pautas, no del análisis.

Al igual que Foster, supongo desde luego que un traductor verterá el discurso indirecto del inglés al francés en la forma usual. A mi modo de ver, lo hará refiriendo una nueva emisión que tendrá que proveer. (Lo mismo ocurre si yo emito "Galileo dijo *that* la Tierra se mueve" dos veces). Admitir esto no equivale a "construir la versión paratáctica de la *oratio obliqua* como una variante notacional de la versión intensional", como insiste Foster. La notación no tiene nada que ver con ello; tanto el semántico del mundo posible como yo aceptamos la misma notación. Diferimos en el análisis semántico. El punto es que el traductor está enunciando el mismo hecho, no en el sentido que le da Foster, sino en algún sentido más usual que a menudo admite, o incluso requiere, que la traducción cambie la referencia cuando esa referencia es, para usar la frase de Reichenbach, una instancia particular reflexiva.

Volviendo a las perspectivas para un análisis paratáctico de ese problemático "enuncia", debemos primero hacer notar que una palabra ligeramente más apropiada sería "implica". Queremos determinar la semántica para emisiones de oraciones como "La teoría T implica *that* 'La nieve es blanca' es verdadero en castellano si y sólo si la nieve es blanca". Y debe sostenerse que una emisión de esta oración se tratará, a los fines de la teoría semántica, como la emisión de dos oraciones, de las cuales la primera termina en un demostrativo que refiere la segunda emisión. Una implicación (de este tipo) resulta ser así una relación entre una teoría y una emisión del hablante que alega la implicación. ¿Qué es esta relación? Una sugerencia razonable dice que es el producto relativo de la relación de consecuencia lógica entre oraciones y la relación de sinonimia entre oraciones y emisiones (quizá de otro lenguaje). Si una teoría T implica que "La nieve es blanca" es verdadero en castellano si y sólo si la nieve es blanca, luego T tiene como consecuencia lógica una oración sinónima de mi emisión de " 'La nieve es blanca' es verdadero en castellano si y sólo si la nieve es blanca".

¿No introduce el segundo componente una apelación a una noción específicamente lingüística: la de sinonimia? Ciertamente: es justo el concepto de traducción que hemos estado tratando de obtener al imponer condiciones a una teoría de la verdad. Esto no resulta en una explicación circular, pues esas condiciones fueron enunciadas (hemos estado suponiendo) en apelar a nociones lingüísticas del tipo que queremos explicar. Así es que el concepto de sinonimia o traducción que yace oculto en la noción de implicación puede usarse sin circularidad cuando nos abocamos a determinar lo que un intérprete conoce. En efecto, cuando atribuimos a un intérprete el concepto de una teoría traduccional ya hemos supuesto ello.

Creo que Foster tiene razón en un aspecto de cierta importancia.

Aun si todo lo que he dicho en defensa de mi formulación de lo que basta para la interpretación es correcto, resta el hecho de que nada constituye estrictamente una teoría del significado. Una teoría de la verdad, no importa cuán bien seleccionada, no es una teoría del significado, mientras que la afirmación de que una teoría traduccional implica ciertos hechos no constituye, debido a los elementos deícticos irreductibles presentes en las oraciones que la expresan, una teoría en el sentido formal. Esto, empero, no torna imposible decir qué es lo que conoce un intérprete, ni tampoco, por ello, dar una respuesta satisfactoria a uno de los problemas principales de la filosofía del lenguaje.

Lenguaje y realidad

13

De la idea misma de un esquema conceptual

Gran número de filósofos de muchas vertientes tienden a hablar de esquemas conceptuales. Los esquemas conceptuales, nos dicen, son formas de organizar la experiencia; son sistemas de categorías que dan forma a los datos de las sensaciones; son puntos de vista desde los cuales individuos, culturas o períodos examinan los acontecimientos que se suceden. Puede no haber traducción de un esquema a otro, en cuyo caso las creencias, deseos, esperanzas y porciones de conocimiento que caracterizan a una persona no tienen contrapartes verdaderas para quien suscribe otro esquema. La realidad misma es relativa a un esquema: lo que cuenta como real en un sistema puede no hacerlo en otro.

Aun aquellos pensadores que tienen la certeza de que hay un solo esquema conceptual se hallan bajo el influjo del concepto de esquemas; aun los monoteístas tienen religión. Y cuando alguien se pone a describir "nuestro esquema conceptual", su doméstica tarea supone, si lo tomamos literalmente, que podría haber sistemas rivales.

El relativismo conceptual es una doctrina temeraria y exótica, o lo sería si pudiéramos comprenderla bien. El problema es, como ocurre tan a menudo en filosofía, que resulta difícil mejorar la inteligibilidad manteniendo simultáneamente la atracción. Sea como fuere, esto es lo que discutiré.

Nos vemos inducidos a imaginar que comprendemos un cambio conceptual generalizado o profundos contrastes por medio de ejemplos legítimos de índole familiar. A veces una idea, como la de simultaneidad que define la teoría de la relatividad, es tan importante que con su agregado toda una sección de la ciencia adquiere un nuevo aspecto. A veces las revisiones en la lista de oraciones consideradas verdaderas en una disciplina son tan esenciales que podemos sentir que los términos involucrados han modificado sus significados. Lenguajes que han evolucionado en tiempos o lugares distantes pueden diferir extensivamente en sus recursos para tratar con uno u otro rango de fenómenos. Lo que surge fácilmente en un lenguaje puede surgir con dificultad en otro, y esta diferencia puede reflejar significativas disimilitudes en estilo y valor.

Pero los ejemplos como éstos, por impactantes que sean en ocasiones, no son tan extremos como para que los cambios y los contrastes no puedan explicarse y describirse usando el equipamiento de un único lenguaje. Whorf, cuando quiere demostrar que el hopi incorpora una metafísica tan extraña a nosotros que el hopi y el inglés, dice él, no pueden

"calibrarse", usa el inglés para conducir los contenidos de los ejemplos de oraciones en hopi.[1] Khun expresa de una manera brillante cómo eran las cosas antes de la revolución usando —¿de qué otro modo podría ser?— nuestro idioma posrevolucionario.[2] Quine nos ofrece una sensación de la "fase preindividuativa en la evolución de nuestro esquema conceptual",[3] en tanto Bergson nos dice adónde podemos ir para disfrutar de un paisaje de montaña que no sufra la distorsión de una u otra perspectiva provincial.

La metáfora dominante del relativismo conceptual, aquella de los puntos de vista diferenciados, parece poner al descubierto una paradoja subyacente. Tiene sentido hablar de distintos puntos de vista, pero sólo si existe un sistema coordinado común en el cual representarlos; sin embargo, la existencia de un sistema común contradice la pretensión de una incomparabilidad profunda. Lo que necesitamos, me parece, es alguna idea de las consideraciones que fijan los límites al contraste conceptual. Hay suposiciones extremas que caen en la paradoja o la contradicción; hay ejemplos modestos que comprendemos sin inconvenientes. ¿Qué determina el límite entre lo meramente raro o novel y lo absurdo?

Podemos aceptar la doctrina que asocia el tener un lenguaje con el tener un esquema conceptual. Puede suponerse que la relación es así: cuando los esquemas conceptuales difieren, también lo hacen los lenguajes. Pero hablantes de diferentes lenguajes pueden compartir un esquema conceptual siempre que haya una vía para traducir un lenguaje al otro. El estudio de los criterios de traducción es por lo tanto una forma de poner el acento en los criterios de identidad de esquemas conceptuales. Si los esquemas conceptuales no se hallan asociados a los lenguajes de este modo, el problema original se ve duplicado innecesariamente, pues entonces tendríamos que imaginar a la mente, con sus categorías ordinarias, operando con un lenguaje con *su* estructura organizativa. Bajo tales circunstancias sin duda querríamos preguntar quién es el que manda.

Como alternativa, está la idea de que *todo* lenguaje distorsiona la realidad, lo cual implica que a lo sumo la mente puede aprehender las cosas tal como realmente son sólo sin usar palabras. Esto equivale a concebir el lenguaje como un medio inerte (si bien necesariamente distorsionador) independiente de los agentes humanos que lo emplean; una visión del lenguaje que seguramente no puede sostenerse. Mas si la mente puede aprehender sin distorsión de lo real, ella misma debe estar desprovista de categorías y conceptos. Esta caracterización sin rasgos propios es común a teorías de segmentos muy diferentes del espectro filosófico. Por ejemplo, hay teorías para las cuales la libertad consiste en decisiones tomadas al margen de todos los deseos, hábitos y disposiciones del agente; y hay teorías del conocimiento que sugieren que la men-

[1] B. L. Whorf, "The Punctual and Segmentative Aspects of Verbs in Hopi".
[2] T. S. Kuhn, *The Structure of Scientific Revolutions.*
[3] W. V. Quine, "Speaking of Objects", 24.

te puede observar la totalidad de sus propias percepciones e ideas. En cada caso, la mente se halla divorciada de los rasgos que la constituyen; una conclusión ineludible a partir de ciertas líneas de razonamiento, como dije antes, pero que siempre debería persuadirnos de rechazar las premisas.

Podemos entonces identificar los esquemas conceptuales con los lenguajes, o mejor, admitiendo la posibilidad de que más de un lenguaje pueda expresar el mismo esquema, con conjuntos de lenguajes intertraducibles. No vamos a considerar que los lenguajes son separables de las almas; un hombre no puede perder la condición de hablar un lenguaje y retener al mismo tiempo el poder del pensamiento. Por lo tanto no es posible que alguien pueda ocupar un punto de observación para comparar esquemas conceptuales desprendiéndose temporariamente de los suyos propios. ¿Podemos decir entonces que dos personas tienen esquemas conceptuales diferentes si hablan lenguajes que carecen de intertraducibilidad?

A continuación considero dos tipos de caso que podría esperarse que aparecieran: los fallos completo y parcial de traducibilidad. Se daría un fallo completo si ninguna clase significativa de oraciones en un lenguaje pudiera ser traducida al otro; el fallo sería parcial si alguna clase pudiera traducirse y alguna clase no (dejaré de lado posibles asimetrías). Mi estrategia consistirá en argumentar que no podemos entender un fallo total, y luego examinaré más brevemente casos de fallo parcial.

Primero, entonces, los susodichos casos de fallo completo. Ciertamente resulta tentador asumir una postura terminante: es dable decir que nada podría considerarse evidencia de que alguna forma de actividad no puede ser interpretada en nuestro lenguaje sin ser al mismo tiempo evidencia de que esa forma de actividad no puede ser una conducta de habla. Si esto fuera correcto, probablemente nos veríamos obligados a sostener que una forma de actividad no puede interpretarse como lenguaje en nuestro lenguaje no es conducta de habla. Mas esto resulta satisfactorio, pues no va más allá de instituir la traducibilidad a una lengua familiar como criterio para la condición de lenguaje. Además, la tesis carece de la atracción que genera la autoevidencia; si es una verdad, como yo pienso que lo es, debería emerger como la conclusión de un argumento.

La credibilidad de esta posición se beneficia al verse reflejada en las estrechas relaciones entre el lenguaje y la atribución de actitudes como creencia, deseo e intención. Por un lado, está claro que el habla reqiere una multitud de intenciones y creencias sutilmente discriminadas. Una persona que afirma que la perseverancia mantiene alto el honor debe, por ejemplo, representar que ella misma cree que la perseverancia mantiene alto el honor, y debe tratar de representarse a sí misma creyéndolo. Por otro lado, parece improbable que podamos atribuir inteligiblemente a un hablante actitudes tan complejas como éstas, a menos que podamos traducir sus palabras a las nuestras. No puede haber dudas de que la relación entre la capacidad de traducir el lenguaje de alguien y la capacidad de describir sus actitudes es muy cercana. Pero,

hasta tanto podamos decir más acerca de *qué* es esta relación, la cuestión en contra de los lenguajes intraducibles permanecerá oscura.

A veces se piensa que la traducibilidad a un lenguaje familiar, digamos al castellano, no puede ser un criterio de condición de lenguaje en tanto la relación de traducibilidad no es transitiva. La idea es que algún lenguaje, digamos el saturniano, puede ser traducible al castellano, y algún otro lenguaje, como el plutoniano, puede ser traducible al saturniano, pero no es traducible al castellano. Una cantidad suficiente de diferencias traducibles pueden sumarse hasta llegar a una intraducible. Si imaginamos una secuencia de lenguajes, cada uno de ellos lo suficientemente cercano al previo como para ser traducido aceptablemente a él, podemos imaginar un lenguaje tan diferente del castellano como para resistir totalmente su traducción a él. En correspondencia a este distante lenguaje habría un sistema de conceptos completamente ajeno a nosotros.

Pienso que este ejercicio no introduce ningún elemento nuevo a la discusión. Porque tendríamos que preguntar cómo reconocimos que lo que el saturniano estaba haciendo era *traducir* el plutoniano (o cualquier otra cosa). El hablante saturniano podría decirnos que eso era lo que él estaba haciendo o más bien podríamos suponer por un momento que nos estaba diciendo eso. Pero luego nos pondríamos a pensar si nuestra traducción del saturniano fue correcta.

Según Kuhn, los científicos que operan en diferentes tradiciones científicas (dentro del diferentes "paradigmas") "trabajan en diferentes mundos".[4] La obra *The Bounds of Sense* de Strawson comienza con la afirmación de que "Es posible imaginar clases de mundos muy diferentes del mundo tal como lo conocemos nosotros".[5] Puesto que hay cuando mucho un mundo, estas pluralidades son metafóricas o meramente imaginadas. No obstante, estas metáforas no son em modo alguno iguales. Strawson nos invita a imaginar posibles mundos no reales, mundos que podrían describirse, usando nuestro lenguaje presente, mediante la redistribución de los valores de verdad para las oraciones según varias formas sistemáticas. La claridad de los contrastes entre los mundos depende en este caso de suponer que nuestro esquema de conceptos, nuestros recursos descriptivos, permanecen fijos. Kuhn, por otra parte, nos dice que pensemos en diferentes observadores del mismo mundo que llegan a él con inconmensurables sistemas de concepto. Los muchos mundos imaginados de Strawson se ven o se escuchan o se describen desde el mismo punto de vista; el mundo único de Kuhn se ve desde diferentes puntos de vista. Esta es la segunda metáfora sobre la cual queremos trabajar.

La primera metáfora requiere una distinción dentro del lenguaje entre concepto y contenido: al usar un sistema fijo de conceptos (palabras con significados fijos) describimos universos alternativos. Algunas

[4] T. S. Kuhn, *The Structure of Scientific Revolutions*, 134.
[5] P. Strawson, *The Bound of Sense*, 15.

oraciones serán verdaderas simplemente por causa de los conceptos o significados involucrados, otras por causa de los rasgos del mundo. Al describir mundos posibles, sólo jugamos con oraciones del segundo tipo.

La segunda metáfora sugiere en cambio un dualismo de un tipo muy diferente, un dualismo de esquema total (o lenguaje) y de contenido ininterpretado. La adhesión al segundo dualismo, si bien no es inconsistente con una adhesión al primero, puede verse estimulada mediante ataques al primero. He aquí cómo puede operar esto.

Dejar de lado la distinción analítico-sintética como distinción básica para la comprensión del lenguaje equivale a dejar de lado la idea de que podemos distinguir claramente entre teoría y lenguaje. El significado, tal como usaríamos vagamente la palabra, está contaminado por la teoría, por lo que se considera verdadero. Feyerabend lo expresa de la siguiente manera:

> Nuestro argumento contra la invariación del significado es simple y claro. Surge del hecho de que usualmente algunos de los principios involucrados en las determinaciones de los significados de las teorías o puntos de vista más viejos son inconsistentes con las nuevas... teorías. El argumento señala que es natural resolver esta contradicción eliminando los viejos principios... problemáticos, para reemplazarlos por principios, o teoremas, de una nueva... teoría. Y concluye mostrando que un procedimiento así llevará también a la eliminación de los viejos significados.[6]

Puede parecer que ahora tenemos una fórmula para generar esquemas conceptuales precisos. Obtenemos un esquema nuevo a partir de uno viejo cuando los hablantes de un lenguaje pasan a aceptar como verdadero un importante rango de oraciones que previamente tomaban por falso (y viceversa, por supuesto). No debemos describir este cambio simplemente como un pasaje en el cual ellos pasan a considerar viejas falsedades como verdades, pues una verdad es una proposición, y lo que ellos pasan a aceptar, al aceptar una oración como verdadera, no es la misma cosa que ellos habían rechazado cuando antes consideraban que la oración era falsa. Se ha operado un cambio sobre el significado de la oración pues ella pertenece ahora a un nuevo lenguaje.

Esta ilustración de la forma es que resultan nuevos (y quizá mejores) esquemas a partir de una ciencia nueva y mejor es muy parecida a la ilustración que los filósofos de la ciencia, como Putnam y Feyerabend, y los historiadores de la ciencia, como Kuhn, han pintado para nosotros. Emerge una idea parecida de la sugerencia de algunos otros filósofos, en el sentido de que podríamos perfeccionar nuestro bagaje conceptual si adecuáramos nuestros lenguaje a una ciencia perfeccionada. Es así que tanto Quine como Smart, en formas algo diferentes, admiten con desagrado que nuestras formas actuales de hablar tornan imposible una ciencia seria de la conducta. (Wittgenstein y Ryle han dicho cosas simi-

[6] P. Feyerabend, "Explanation, Reduction, and Empiricism", 82.

lares sin lamentar la situación.) La cura, piensa Quine y Smart, consiste en cambiar la forma en que hablamos. Smart aboga por (y predice) el cambio de manera de colocarnos en el camino científicamente recto del materialismo: Quine está más interesado en despejar la vía hacia un lenguaje puramente extensional. (Tal vez yo debería agregar que pienso que nuestro esquema y nuestro lenguaje actuales se entienden mejor como extensionales y materialistas.)

Si fuéramos a seguir este consejo, personalmente no pienso que la ciencia o la comprensión avanzarían, aunque es posible que la moral sí lo hiciera. Pero la cuestión presente es sólo si, para el caso de que tales cambios fueran a tener lugar, se justificaría que los llamáramos alteraciones en el aparato conceptual básico. La dificultad de llamarlos así es fácil de apreciar. Supongamos que desde mi oficina de Ministro del Lenguaje Científico quiero que el hombre nuevo deje de usar palabras que refieran, digamos, emociones, sentimientos, pensamientos e intenciones, y hable en cambio de los estados y sucesos fisiológicos que se supone son más o menos idénticos al ajetreo mental. ¿Cómo sé si mi consejo ha sido tenido en cuenta, dado que el hombre nuevo habla un nuevo lenguaje? Por cuanto yo sé, las relucientes frases nuevas, si bien han sido extraídas del viejo lenguaje en el cual refieren movimientos fisiológicos, pueden en su boca desempeñar el papel de los confusos conceptos mentales viejos.

La frase clave es: por cuanto yo sé. Está claro que la retención de parte o de todo el viejo vocabulario no proporciona en sí misma una base para juzgar si el nuevo esquema es igual al viejo o diferente de él. Así, lo que al principio sonaba como un descubrimiento estremecedor —que la verdad es relativa a un esquema conceptual— hasta ahora no ha podido demostrarse que sea algo más que el hecho pedestre y familiar de que la verdad de una oración es relativa (entre otras cosas) al lenguaje al cual ella pertenece. En vez de vivir en mundo diferentes, los científicos de Kuhn pueden estar, como quienes necesitan el diccionario Webster's, separados solamente por palabras.

El dejar de lado la distinción analítico-sintética no ha probado ser de ayuda para comprender el relativismo conceptual. La distinción analítico-sintética se ve explicada no obstante en términos de algo que puede servir para apuntalar al relativismo conceptual, esto es, la idea de contenido empírico. El dualismo de lo sintético y lo analítico es un dualismo de oraciones, algunas de las cuales son verdaderas (o falsas) tanto por causa de lo que significan como por causa de su contenido empírico, mientras que otras son verdaderas (o falsas) sólo en virtud de su significado, careciendo de contenido empírico. Si descartamos el dualismo, abandonamos la concepción de significado que él conlleva, pero no tenemos que abandonar la idea de contenido empírico: podemos sostener, si queremos, que *todas* las oraciones tienen contenido empírico. El contenido empírico se explica a su vez por referencia a los hechos, al mundo, a la experiencia, a la sensación, a la totalidad de estímulos sensoriales, o a algo similar. Los significados nos dieron una vía para hablar acerca de las categorías, de la estructura organizadora del lenguaje, y así por el

estilo; pero es posible, como hemos visto, desechar los significados y la analiticidad mientras se conserva la idea de que el lenguaje engloba un esquema conceptual. Así, en lugar del dualismo de lo analítico-sintético tenemos el dualismo de esquema conceptual y contenido empírico. El nuevo dualismo crea los cimientos de un empirismo al que se han amputado los dogmas insostenibles de la distinción analítico-sintética y del reduccionismo, aliviándolo, así, de la idea impracticable de que podemos asignar un contenido empírico individualmente, oración por oración.

Quiero manifestar que este segundo dualismo de esquema y contenido, de un sistema organizador y de algo que espera ser organizado, no puede estatuirse como inteligible y defendible. Es en sí mismo un dogma del empirismo, el tercer dogma. El tercero, y quizás el último, puesto que si lo descartamos no resulta claro que vaya a quedar algo característico que pueda llamarse empirismo.

El dualismo esquema-contenido ha sido formulado de diferentes maneras. He aquí algunos ejemplos. El primero de ellos proviene de Whorf, quien trabaja sobre un tema de Sapir. Whorf dice que:

> ...el lenguaje produce una organización de la experiencia. Nos inclinamos por pensar que el lenguaje es simplemente una técnica de expresión, y no por entender que el lenguaje es ante todo una clasificación y organización del flujo de experiencia sensorial que resulta en un cierto orden del mundo... En otras palabras, el lenguaje hace en una forma más cruda pero también más amplia y versátil lo mismo que hace la ciencia... Se nos presenta así un nuevo principio de relatividad, que sostiene que los observadores no están todos guiados por la misma evidencia física hacia una misma ilustración del universo, a menos que sus encuadres lingüísticos sean similares, o puedan ser calibrados de alguna manera.[7]

Tenemos aquí todos los elementos que se requiere: el lenguaje como la fuerza organizadora, que no se distingue claramente de la ciencia; lo que es organizado, referido de varias maneras como "experiencia", "el flujo de experiencia sensorial", y "evidencia física", y finalmente, el fallo de intertraducibilidad ("calibrador"). El fallo de intertraducibilidad es una condición necesaria para diferenciar los esquemas conceptuales; se supone que lo que nos ayuda a comprender la afirmación de que cuando la traducción falla los que están bajo consideración son lenguajes o esquemas es la relación común con la experiencia o la evidencia. Es esencial para esta idea que haya algo neutral y común situado fuera de todos los esquemas. Desde luego, este algo común no puede ser la *materia* de los lenguajes contrastantes, o la traducción sería posible. Es así como Kuhn ha escrito recientemente:

> Lo filósofos han abandonado hoy la esperanza de encontrar un lenguaje de puros *sense-data*... pero muchos de ellos continúan suponiendo que las

[7] B. L. Whorf, "The Punctual and Segmentative Aspectos of Verbs in Hopi", 55.

teorías pueden compararse mediante el recurso a un vocabulario básico que consista sólo en palabras vinculadas a la naturaleza de maneras no problemáticas y, en el grado necesario, independientes de la teoría... Feyerabend y yo hemos argumentado ampliamente que un tal vocabulario no es accesible. En la transición de una teoría a la siguiente las palabras cambian sus significados o condiciones de aplicabilidad de un modo sutil. Si bien los mismos signos se siguen usando en su mayor parte antes y después de una revolución —por ejemplo, fuerza, masa, elemento, compuesto, célula— la forma en que algunos de ellos se vinculan a la naturaleza cambia en alguna medida. Decimos por esto que teorías sucesivas son inconmensurables.[8]

"Inconmensurable" es, por supuesto, la palabra de Kuhn y Feyerabend para decir "no intertraducible". El contenido neutral que espera ser organizado lo proporciona la naturaleza.

El propio Feyerabend sugiere que podemos comparar esquemas contrastantes mediante "la elección de un punto de vista que esté fuera del sistema o del lenguaje". Espera que podamos hacerlo pues "aún hay experiencia humana como proceso realmente existente"[9] independiente de todos los esquemas

Quine expresa pensamientos iguales o similares en muchos pasajes: "La totalidad de lo que damos en llamar conocimiento o creencias... es un tejido hecho por el hombre que tropieza con la experiencia sólo a lo largo de sus bordes...";[10] "...la ciencia en su totalidad es como un campo de fuerza cuyas condiciones límite son la experiencia";[11] "Como empirista...pienso que el esquema conceptual de la ciencia es una herramienta...para predecir la experiencia futura a la luz de la experiencia pasada."[12] Y nuevamente:

Insistimos en descomponer de alguna manera la realidad en una multiplicidad de objetos identificables y discriminables... Hablamos tan inveteradamente de los objetos que decir que lo hacemos casi parece como no decir nada; pues ¿hay otra manera de hablar? Es difícil decir de qué otra manera se puede hablar, no porque nuestro patrón de objetivación sea un rasgo invariable de la naturaleza humana, sino porque estamos constreñidos a adaptar a nuestro patrón propio todo patrón extraño en el proceso mismo de comprensión o traducción de las oraciones extrañas.[13]

El test de diferencia continúa siendo el fallo o la dificultad de traducción: "...decir que ese remoto medio es radicalmente diferente del nuestro no es decir que las traducciones no se logran fácilmente."[14] Y la

8 T. S. Kuhn, "Reflections on my Critics", 266, 267.
9 P. Feyerabend, "Problems of Empiricism", 214.
10 W. V. Quine, "Two Dogmas of Empiricism", 42.
11 *Ibídem.*
12 *Ibídem*, 44.
13 W. V. Quine, "Speaking of Objets", 1.
14 *Ibídem*, 25.

dificultad puede ser tan grande que decimos que el extraño emplea "un patrón más allá de toda individuación, hasta ahora inimaginado".[15]

La idea es entonces que algo es un lenguaje, y está asociado a un esquema conceptual, podamos o no traducirlo, si se encuentra en una cierta relación (de predicción, de organización, de estudio, o de ajuste) con la experiencia (naturaleza, realidad, apuntes sensoriales). El problema es decir qué es la relación, y dar una idea más clara de las entidades relacionadas.

Las imágenes y metáforas se reúnen en dos grupos principales: los esquemas conceptuales (lenguajes) u *organizan* algo, o lo *ajustan* (como en "reorienta su herencia científica para ajustarla a sus incitaciones sensoriales").[16] El primer grupo incluye también *sistematizar*, *distribuir* (el flujo de experiencia); otros ejemplos del segundo grupo son *predecir*, *explicar*, *enfrentar* (el tribunal de la experiencia). En cuanto a las entidades que son organizadas, o que el esquema deben ajustar, pienso nuevamente que podemos detectar dos ideas principales: o es la realidad (el universo, el mundo, la naturaleza) o es la experiencia (los sucesos que se nos manifiestan, las irritaciones de superficie, las incitaciones sensoriales, los *sense-data*, lo dado).

No podemos asignar un significado claro a la noción de organizar un único objeto (el mundo, la naturaleza, etcétera) a menos que se entienda que el objeto contiene o consiste en otros objetos. Alguien que se pone a organizar un armario ordena las cosas que hay en su interior. Si a usted le dijeran que no organizara los zapatos y las camisas, sino el propio armario, usted se quedaría perplejo. ¿Cómo organizaría usted el Océano Pacífico? Enderezaría sus costas, tal vez, o reubicaría sus islas, o destruiría sus peces.

Un lenguaje puede contener predicados simples cuyas extensiones no son correspondidas por predicados simples, o incluso por ningún predicado, en algún otro idioma. Lo que nos permite lograr este objetivo en casos particulares es una ontología común a los dos lenguajes, con conceptos que individuan los mismos objetos. Podemos tener claros los fracasos de traducción cuando éstos son lo suficientemente locales, pues un marco de traducción generalmente exitosa proporciona lo necesario para hacer inteligibles los fallos. Pero nuestra meta era aun mayor: queríamos dar sentido a la existencia de un lenguaje que no pudiéramos traducir en absoluto. O, para expresarlo de otro modo, estábamos buscando un criterio de condición de lenguaje que no dependiera de la traducibilidad a un idioma familiar, o que no la implicara. Presumo que la imagen de organizar el armario de la naturaleza no proporcionará ese criterio.

¿Qué hay del otro tipo de objeto, la experiencia? ¿Podemos pensar en un lenguaje que *la* organice? Vuelven a surgir dificultades parecidas. La noción de organización sólo es aplicable a pluralidades. Pero cualquiera que sea la pluralidad en que a nuestro entender consista la expe-

[15] *Ibídem*, 24.
[16] W. V. Quine, "Two Dogmas of Empiricism", 46.

riencia —sucesos como perder un botón o lastimarse un dedo, tener una sensación de calor o escuchar un oboe— tendremos que individuar según principios familiares. Un lenguaje que organice *esas* entidades debe ser un lenguaje muy parecido al nuestro.

La experiencia (y sus colegas como las irritaciones de superficie, las sensaciones y los *sense-data*) pone también en una dificultad todavía más obvia a la idea organizadora. Pues ¿cómo podría llamarse lenguaje a algo que organizara *sólo* experiencias, sensaciones, irritaciones de superficie o *sense-data*? Seguramente los cuchillos y los tenedores, los ferrocarriles y las montañas, los repollos y las monarquías también necesitan organización.

Esta última observación sonará sin duda inapropiada como respuesta a la pretensión de que un esquema conceptual es una manera de habérselas con la experiencia sensorial; y estoy de acuerdo en que es así. Pero lo que estaba bajo consideración era la idea de *organizar* la experiencia, no la idea de *habérselas con* (o ajustar o encarar) la experiencia. La respuesta se dio a propósito del primer concepto, no del último. Entonces veamos ahora si podemos llegar a algo mejor con la segunda idea.

Cuando pasamos de hablar de organización a hablar de ajuste orientamos nuestra atención del aparato referencial del lenguaje —predicados, cuantificadores, variables y términos singulares— a las oraciones completas. Son las oraciones las que predicen (o se usan para predecir), las que hacen frente a las cosas o tratan con ellas, las que ajustan nuestras incitaciones sensoriales, las que pueden compararse o confrontarse con la evidencia. Son las oraciones también, las que enfrentan al tribunal de la experiencia, aunque por supuesto deben enfrentarlo juntas.

La propuesta no dice que las experiencias, los *sense-data*, las irritaciones de superficie o las incitaciones sensoriales sean la única materia del lenguaje. Allí está, es cierto, la teoría que sostiene que el habla acerca de las casas de ladrillos de Elm Street se construye en última instancia como si tratara de *sense-data* o de percepciones, pero tales visiones reduccionistas son tan sólo versiones extremas e inadmisibles de la posición general que estamos considerando. Esta posición general sostiene que la experiencia sensorial provee toda la *evidencia* para la aceptación de las oraciones (donde las oraciones pueden incluir teorías completas). Una oración o teoría se ajusta a nuestras incitaciones sensoriales, enfrenta con éxito al tribunal de la experiencia, predice la experiencia futura, o hace frente a los patrones de nuestras irritaciones de superficie, siempre que esté confirmada por la evidencia.

En el curso normal de los acontecimientos, una teoría puede estar confirmada por la evidencia disponible y sin embargo ser falsa. Pero lo que aquí tenemos en vista no es sólo la evidencia realmente disponible; es la totalidad de la evidencia sensorial posible pasada, presente y futura. No necesitamos detenernos a contemplar lo que esto podría significar. La cuestión es: que una teoría se ajuste o enfrente a la totalidad de la evidencia sensorial posible equivale a que esa teoría sea verdadera. Si una teoría cuantifica sobre objetos físicos, números o conjuntos, lo

que ella dice acerca de estas entidades es verdadero en tanto la teoría como un todo se ajuste a la evidencia sensorial. Uno puede ver que, desde este punto de vista, dichas entidades pueden ser llamadas supuestos. Es razonable decir que algo es un supuesto si se lo puede contrastar con algo que lo es. Aquí, el "algo que no lo es" es la experiencia sensorial; al menos ésa es la idea.

El problema es que la noción de ajustarse a la totalidad de la experiencia, como la noción de ajustarse a los hechos, o de ser fiel a los hechos, no agrega nada inteligible al simple concepto de ser verdadero. Hablar de experiencia sensorial en vez de hablar de evidencia, o simplemente de los hechos, expresa una visión acerca de la fuente o naturaleza de la evidencia, pero no suma una nueva entidad al universo respecto de la cual puedan ponerse a prueba los esquemas conceptuales. La totalidad de la evidencia sensorial es aquello que requerimos siempre y cuando sea toda la evidencia que exista; y es justo toda la evidencia existente lo que hace falta para que nuestras oraciones o teorías sean verdaderas. Sin embargo, nada, ninguna *cosa*, hace verdaderas las oraciones y las teorías: ni la experiencia, ni las irritaciones de superficie, ni el mundo pueden hacer verdadera una oración. Que la experiencia toma un rumbo determinado, que nuestra piel es tibia o está lastimada, que el universo es finito, estos hechos, si queremos decirlo de esta manera, hacen que las oraciones y las teorías sean verdaderas. Pero esto se ve mejor sin hacer mención de los hechos. La oración "Mi piel es tibia" es verdadera si y sólo si mi piel es tibia. Aquí no hay referencia alguna a un hecho, un mundo, una experiencia, o una porción de evidencia.[17]

Nuestro intento de caracterizar los lenguajes o los esquemas conceptuales en términos de la noción de ajustarse a alguna entidad se ha reducido, entonces, al simple pensamiento de que algo es un esquema conceptual o teoría aceptable si es verdadero. Quizá sea mejor *mayormente* verdadero de manera de admitir que quienes comparten un esquema difieran en los detalles. Y el criterio de un esquema conceptual diferente del nuestro se convierte ahora en: mayormente verdadero pero no traducible. La cuestión de que este sea o no un criterio útil es sólo una cuestión de cuán bien comprendamos la noción de verdad, aplicada al lenguaje, independientemente de la noción de traducción. La respuesta es, pienso, que no la comprendemos en absoluto en forma independiente.

Reconocemos que oraciones como " 'La nieve es blanca' es verdadero si y sólo si la nieve es blanca" son trivialmente verdaderas. Sin embargo la totalidad de tales oraciones castellanas determina singularmente la extensión del concepto de verdad para el castellano. Tarski generalizó esta observación y la convirtió en un test de las teorías de la verdad: de acuerdo con la Convención T de Tarski, una teoría satisfactoria de la verdad para un lenguaje L debe implicar, para cada oración *s* de L, un teorema de la forma "*s* es verdadero si y sólo si p" donde "*s*" es reempla-

[17] Véase el Ensayo 3.

zado por una descripción de *s* y "*p*" por *s* mismo si L es castellano, y por una traducción de *s* al castellano si L no es castellano.[18] Esto no es, desde luego, una definición de la verdad, ni tampoco insinúa que haya una única definición o teoría que sea aplicable a los lenguajes en general. No obstante, la Convención T sugiere, si bien no puede enunciar, una característica importante común a todos los conceptos especializados de la verdad. Este logro se debe al uso esencial que da a la noción de traducción a un lenguaje que conocemos. Puesto que la Convención T encarna nuestra mejor intuición de la forma en que se usa el concepto de verdad, no parece haber mucho futuro para un test que busque probar que un esquema conceptual puede ser radicalmente diferente del nuestro si ese test depende de la presunción de que podemos separar la noción de verdad de la de traducción.

Ni un repertorio fijo de significado, ni una realidad que se mantenga neutral frente a las teorías, pueden proporcionar, entonces, una base para la comparación de esquemas conceptuales. Sería un error ir más allá en la búsqueda de dicha base si con ello queremos significar un algo común a esquemas inconmensurables. Al abandonar esta búsqueda, abandonamos el intento de comprender la metáfora de un espacio único dentro del cual cada esquema tiene una posición y provee un punto de vista.

Me ocuparé ahora de la aproximación más modesta: la idea de fallo parcial, en vez de total, de la traducción. Esto introduce la posibilidad de hacer cambios y contraposiciones en los esquemas conceptuales, inteligibles mediante referencia a la parte común. Lo que necesitamos es una teoría de la traducción o de la interpretación que no efectúe suposiciones acerca de significados, conceptos o creencias compartidos.

La interdependencia de creencia y significado nace de la interdependencia de dos aspectos de la interpretación de la conducta de habla: la atribución de creencias y la interpretación de oraciones. Señalamos antes que a causa de estas dependencias podemos permitirnos asociar esquemas conceptuales con lenguajes. Ahora podemos expresar el asunto de una manera algo más definida. Admitamos que el habla de un hombre sólo puede ser interpretado por alguien que conozca bastante acerca de lo que el hablante cree (y pretende y quiere), y que las distinciones finas entre creencias son imposibles sin la comprensión del habla; ¿cómo vamos entonces a interpretar el habla o a atribuir inteligiblemente creencias y otras actitudes? Está claro que debemos tener una teoría que simultáneamente tenga en cuenta actitudes e interprete el habla, y que no suponga ninguno de ambos.

Sugiero, siguiendo el ejemplo de Quine, que podemos aceptar, sin caer en circularidades ni presunciones sin garantías, ciertas actitudes muy generales hacia las oraciones como evidencia básica para una teoría de la interpretación radical. En beneficio de la presente discusión al menos podemos depender de la actitud de aceptar como verdadera, diri-

[18] A Tarski, "The Concept of Truth in Formalized Languages".

gida a las oraciones, como noción crucial. (Un teoría más audaz tomaría también otras actitudes hacia las oraciones, tales como desear que sea verdadera, imaginarse si es verdadera, pretender hacer verdadera, y otras por el estilo.) Por cierto, las actitudes se encuentran involucradas aquí, pero el hecho de que no se deje de lado el asunto central puede verse en esto: si nosotros meramente sabemos que alguien considera que una determinada oración es verdadera, no sabemos ni lo que ese alguien significa con la oración ni qué creencia representa el hecho de que la considere verdadera. Su consideración de la oración como verdadera es así el vector de dos fuerzas: el problema de la interpretación es abstraer de la evidencia una teoría viable del significado y una teoría aceptable de la creencia.

La manera en que se resuelve este problema se aprecia mejor partiendo de ejemplos poco llamativos. Si usted ve un yate navegando y su compañero dice, "Mira qué hermoso bote", usted puede enfrentarse a un problema de interpretación. Una posibilidad natural es que su amigo haya confundido un yate con un bote, y se haya formado una creencia falsa. Pero si la vista de él es buena y su punto de observación favorable es aun más plausible que él no use la palabra "bote" tal como usted lo hace, y no haya cometido error alguno respecto de las características de la embarcación que pasaba. Todo el tiempo estamos efectuando este tipo de interpretaciones improvisadas, decidiendo en favor de la reinterpretación de palabras de manera de preservar una razonable teoría de la creencia. Como filósofos tenemos una peculiar tolerancia hacia el malapropismo sistemático, y somos prácticos en la interpretación de sus resultados. El proceso consiste en construir una teoría viable de la creencia y del significado a partir de oraciones consideradas verdaderas.

Esta clase de ejemplos enfatiza la interpretación de detalles anómalos en relación con un marco de creencias comunes y a un método de traducción aplicable. Pero los principios comprometidos deben ser los mismos para casos menos triviales.Lo que interesa es lo siguiente: si todo lo que conocemos son las oraciones que el hablante considera verdaderas, y no podemos suponer que su lenguaje sea el nuestro, luego no podemos avanzar siquiera un primer paso hacia la interpretación sin conocer o suponer una gran parte de las creencias del hablante. Puesto que el conocimiento de las creencias sólo aparece con la capacidad de interpretar palabras, al principio la única posibilidad es suponer un acuerdo general respecto de las creencias. Nuestra primera aproximación a una teoría terminada la tenemos al asignar a las oraciones de un hablante condiciones de verdad que ellas realmente obtienen (en nuestra propia opinión) precisamente cuando el hablante considera que esas oraciones son verdaderas. La política que nos debe guiar es la de aplicar esto en la medida de lo posible, ateniéndonos a consideraciones de simplicidad, presentimientos acerca de los efectos del condicionamiento social, y por supuesto nuestro conocimiento, científico o de sentido común, de error explicable.

El método no está destinado a eliminar los desacuerdos, ni puede hacerlo; su propósito es hacer posible el desacuerdo significativo, y esto

depende enteramente de una fundamentación —*alguna* fundamentación— en el acuerdo. El acuerdo puede tomar la forma de un espectro ampliamente compartido de oraciones consideradas verdaderas por hablantes del "mismo lenguaje", o de un acuerdo mediado a grandes rasgos por una teoría de la verdad ideada por un intérprete para hablantes de otro lenguaje.

Dado que la caridad no es una opción, sino una condición para tener una teoría practicable, carece de sentido sugerir que en caso de respaldarla podríamos caer en un error generalizado. Hasta que hayamos establecido con éxito una correlación sistemática de oraciones consideradas verdaderas con oraciones consideradas verdaderas, no hay errores factibles. La caridad nos es impuesta; nos guste o no, si queremos comprender a los demás, debemos darlos por acertados en la mayor parte de los asuntos. Si somos capaces de producir una teoría que reconcilie la caridad y las condiciones formales para una teoría, hemos hecho todo lo que puede hacerse para asegurar la comunicación. No hay nada más que sea posible, ni hace falta nada más.

Comprendemos al máximo las palabras y pensamientos de otros cuando interpretamos en una forma que optimice el acuerdo (esto incluye un margen, como dijimos antes, para el error explicable, como por ejemplo las diferencias de opinión). ¿Dónde deja esto un espacio para el caso del relativismo conceptual? La respuesta es, pienso yo, que mayormente debemos decir lo mismo de las diferencias de esquema conceptual que lo que decimos de las diferencias de creencia: incrementamos la claridad y el atractivo de las declaraciones de diferencia, sean de esquemas o de opinión, ampliando las bases del lenguaje compartido (traducible) o de la opinión compartida. Por cierto, no puede trazarse una línea clara entre los casos. Si nos decidimos a traducir alguna oración extraña que sus hablantes rechazan por una oración con la cual nos unen fuertes vínculos de índole comunitaria, podríamos vernos tentados de llamar a esto una diferencia de esquemas; si decidimos hacer un espacio a la evidencia por otras vías, puede resultar más natural hablar de una diferencia de opinión. Pero cuando otros piensen de manera diferente a nosotros, ningún principio general ni recurso a la evidencia nos podrá obligar a decidir que la diferencia descansa en nuestras creencias más que en nuestros conceptos.

Pienso que debemos concluir que el intento de dar un significado sólido a la idea de relativismo conceptual, y por tanto a la idea de un esquema conceptual, no obtiene mejores resultados cuando se apoya en el fallo parcial de traducción que cuando se apoya en el fallo total. Dada la metodología de interpretación subyacente no podríamos estar en condiciones de juzgar que otras personas tienen conceptos o creencias radicalmente diferentes de los nuestros.

Sería un error resumir lo dicho hasta ahora diciendo que hemos demostrado cómo es posible la comunicación entre personas que tienen diferentes esquemas, algo que funciona sin necesidad de lo que no puede existir, entendiéndose por esto un terreno neutral, o un sistema coordinado común. Y esto es así porque no hemos hallado ningún funda-

mento inteligible en base al cual pueda decirse que los esquemas son diferentes. Sería igualmente erróneo anunciar la gloriosa novedad de que la humanidad toda —toda los hablantes de un lenguaje, al menos— comparte un esquema y una ontología comunes. Pues si no podemos decir inteligiblemente que los esquemas son diferentes, tampoco podemos decir inteligiblemente que son uno solo.

Al dejar de lado la dependencia del concepto de una realidad ininterpretada, de algo exterior a todos los esquemas y a toda la ciencia, no estamos renunciando a la noción de verdad objetiva: todo lo contrario. Dado el dogma de un dualismo de esquema y relidad, llegamos a la relatividad conceptual y a la verdad relativa a un esquema. Sin este dogma, esta clase de relatividad se derrumba sin remedio. Damos por sentado que la verdad de las oraciones continúa siendo relativa al lenguaje, pero eso es todo lo objetivo que puede llegar a ser. Al dejar de lado el dualismo de esquema y mundo, no dejamos de lado al mundo, sino que restablecemos un contacto sin mediaciones con los objetos familiares cuyas travesuras y extravagancias hace a nuestras oraciones y opiniones verdaderas o falsas.

14

El método de la verdad en metafísica

Al compartir un lenguaje, cualquiera sea el sentido en que esto se requiere para la comunicación, compartimos una ilustración del mundo que, en sus rasgos más amplios, debe ser verdadera. Se sigue que cuando hacemos manifiestos los rasgos más amplios de nuestro lenguaje, estamos haciendo manifiestos los rasgos más amplios de la realidad. Por lo tanto, una manera de ir en pos de la metafísica consiste en estudiar la estructura general de nuestro lenguaje. Este no es, desde luego, el único método verdadero de la metafísica; no existe tal cosa. Pero es un método, y ha sido practicado por filósofos tan distanciados entre sí por el tiempo o por la doctrina como Platón, Aristóteles, Hume, Kant, Russell, Frege, Wittgenstein, Carnap, Quine y Strawson. Estos filósofos, no hace falta decirlo, no se han puesto de acuerdo respecto de cuáles son los rasgos más amplios del lenguaje, o respecto de cómo se los puede estudiar y describir mejor; las conclusiones metafísicas han sido en consecuencia diversas.

El método que describiré y recomendaré no es nuevo; cada rasgo importante del mismo puede hallarse en este o en aquel filósofo, y la idea rectora está implícita en gran parte de las mejores obras de la filosofía del lenguaje. Nuevas son la formulación explícita de la propuesta, y la argumentación de su importancia filosófica. Comienzo por la argumentación; luego sigue una descripción del método; finalmente, se esbozan algunas aplicaciones.

I

¿Por qué nuestro lenguaje —cualquier lenguaje— debe incorporar o depender de una visión mayormente correcta, compartida, de cómo son las cosas? Primero consideremos por qué aquellos que pueden entenderse mutuamente en el habla deben compartir una visión del mundo, sea o no correcta esa visión. La razón es que afectamos la inteligibilidad de nuestras lecturas de las emisiones de otros cuando nuestro método de lectura coloca a los otros dentro de lo que tomamos como amplio error. Podemos dar un sentido correcto a las diferencias, pero sólo dentro de un marco de creencia compartida. En general, lo que se comparte no mueve a comentarios; es demasiado insulso, trivial o familiar como para llamar la atención. Pero sin un vasto terreno común, no hay

un lugar donde los contendientes puedan llevar adelante su discusión. Por supuesto, será más fácil no estar de acuerdo que estarlo con alguien con quien no se comparta mucho, pero esto quizá sea obvio.

Las creencias se identifican y describen sólo dentro de un denso patrón de creencias. Yo puedo creer que una nube está pasando delante del sol, pero solamente porque creo que hay un sol, que las nubes están hechas de vapor de agua, que el agua puede existir en forma líquida o gaseosa; y así por el estilo, interminablemente. No hay una lista particular de otras creencias que se requieran para dar sustancia a mi creencia de que una nube está pasando frente al sol; pero allí debe haber algún conjunto apropiado de creencias relacionadas. Si yo supongo que usted cree que una nube está pasando frente al sol, supongo que usted tiene el tipo correcto de patrón de creencias para sostener esa creencia, y estas creencias que yo presumo que usted tiene deben ser, para poder cumplir con su tarea de apoyo, lo suficientemente parecidas a mis creencias como para justificar la descripción de su creencia como una creencia de que una nube está pasando frente al sol. Si yo estoy en lo cierto cuando atribuyó la creencia a usted, luego usted debe tener un patrón de creencias muy parecido al mío. No es extraño, entonces, que yo pueda interpretar sus palabras correctamente sólo interpretando en una forma tal que nos ponga en un amplio acuerdo.

Puede parecer que hasta ahora el argumento solamente muestra que una buena interpretación engendra consenso, en tanto deja bastante abierta la cuestión de si aquello en torno a lo cual hay acuerdo es verdadero. Y ciertamente, el acuerdo, no importa cuál sea su amplitud, no garantiza la verdad. Pero esta observación escapa al núcleo del argumento. La idea básica consiste en que se necesita una gran comunidad de creencia para proporcionar una base para la comunicación o la comprensión; esta idea, expandida, tendría que ser entonces que el error objetivo puede tener tener lugar sólo en un campo de creencia mayormente verdadera. El acuerdo no hace la verdad, pero mucho de lo acordado debe ser verdadero si es que algo de lo acordado es falso.

Así como un exceso de error atribuido expone al sujeto a quedarse sin su materia, igualmente un exceso de error real priva a una persona de cosas respecto de las cuales equivocarse. Cuando queremos interpretar, trabajamos sobre alguna suposición acerca del patrón general de acuerdo. Suponemos que mucho de lo que consideramos común es verdadero, pero, por supuesto, no podemos suponer que sabemos dónde descansa la verdad. No podemos interpretar en base a las verdades conocidas, no porque no conozcamos ninguna, sino porque no siempre sabemos cuáles son ellas. No necesitamos ser omniscientes para interpretar, pero no hay nada de absurdo en la idea de un intérprete omnisciente; éste atribuye creencias a los otros, e interpreta su habla en base a sus propias creencias, tal como lo hacemos el resto de nosotros. Dado que hace esto de la misma manera que nosotros, por fuerza encuentra tanto acuerdo como haga falta para dar sentido a sus atribuciones e interpretaciones; y en este caso, desde luego, lo que se acuerda es verdadero por hipótesis. Pero ahora está claro por qué hablar de error genera-

lizado acerca del mundo es algo simplemente inteligible, pues suponerlo inteligible es suponer que podría haber un intérprete (el intérprete omnisciente) que interpretara correctamente que alguna otra persona está equivocada en general, y hemos demostrado que esto es imposible.

II

Una comunicación exitosa prueba la existencia de una visión del mundo compartida y mayormente verdadera. Pero lo que nos llevó a exigir una visión común fue el reconocimiento de que las oraciones consideradas verdaderas —los representantes lingüísticos de la creencia— determinan los significados de las palabras que contienen. De esta manera, la visión común da forma al lenguaje compartido. A esto se debe que sea plausible sostener que al estudiar los aspectos más generales del lenguaje estaremos estudiando los aspectos más generales de la realidad. Queda por decir cómo pueden identificarse y describirse estos aspectos.

El lenguaje es un instrumento de comunicación debido a su dimensión semántica, a la potencialidad para la verdad o la falsedad de sus oraciones o, mejor, de sus emisiones e inscripciones. El estudio de cuáles oraciones son verdaderas es en general tarea de las diversas ciencias; pero el estudio de las condiciones de verdad es el dominio de la semántica. Aquello a lo que debemos prestar atención en el lenguaje, si es que queremos poner de relieve las características generales del mundo, es qué significa en general para una oración del lenguaje ser verdadera. La sugerencia dice que si las condiciones de verdad de las oraciones se encuentran dispuestas en el contexto de una teoría de amplios alcances, la estructura lingüística que emerge reflejará rasgos importantes de la realidad.

La meta es una teoría de la verdad para una parte razonablemente poderosa y significativa de un lenguaje natural. El campo de la teoría —qué propoción del lenguaje está captada por la teoría, y cuán convincentemente— será un factor del cual dependerá el interés de cualquier resultado metafísico. La teoría debe mostrarnos cómo podemos observar que cada una de las infinitas oraciones potenciales está compuesta a partir de un repertorio finito de átomos semánticamente significativos (es decir, palabras) por medio de un número finito de aplicaciones de un número finito de reglas de composición. Ella debe dar así las condiciones de verdad para cada oración (en relación con las circunstancias de su emisión) en base a su composición. Puede decirse luego que la teoría explica las condiciones de verdad de una emisión de una oración en base a los roles de las palabras en la oración.

Mucho de esto se lo debemos a Frege. El vio la importancia de describir la forma en que la verdad de una oración depende de las características semánticas de sus partes, y sugirió cómo podría hacerse esa descripción para segmentos muy importantes del lenguaje natural. Su método hoy nos resulta familiar: introdujo una notación estandarizada

cuya sintaxis reflejaba directamente la interpretación que se pretendía, y luego argumentó que la nueva notación, al ser interpretada, tenía el mismo poder expresivo que partes importantes del lenguaje natural. O mejor dicho, no exactamente el mismo poder expresivo, puesto que Frege creía que el lenguaje natural es defectuoso en algunos aspectos, y consideraba a su nuevo lenguaje como un progreso.

A Frege le preocupaban la estructura semántica de las oraciones y las relaciones semánticas entre oraciones, en cuanto éstas generaban implicaciones. Pero no puede decirse que él haya concebido la idea de una amplia teoría formal de la verdad para un lenguaje como un todo. Una consecuencia de ello fue su falta de interés en las paradojas semánticas. Otra consecuencia fue un deseo aparente de aceptar una infinidad de significados (sentidos) y referentes para cada frase denotativa del lenguaje.

Debido a que consideró que la aplicación de función al argumento era el único modo de combinación semántica, Frege se vio obligado a tratar las oraciones como una clase de nombre: el nombre de un valor de verdad. Visto simplemente como un astuto esquive en el camino hacia la caracterización de las condiciones de verdad de las oraciones, este recurso de Frege es intachable. Pero siendo que las oraciones no operan en el lenguaje en forma en que lo hacen los nombres, la aproximación de Frege debilita toda confianza en que la ontología que él necesita para producir su semántica tenga alguna conexión directa con la ontología implícita en el lenguaje natural. No queda claro, entonces, qué puede uno aprender acerca de la metafísica partiendo del método de Frege. (Por cierto, no quiero decir con esto que con el método de Frege no podamos aprender nada de la metafísica; pero para ver cómo hacerlo habrá que reunir argumentos distintos del mío.)

Quine aportó un ingrediente esencial para el proyecto en cuestión cuando mostró la forma en que una aproximación holística al problema de la comprensión de un lenguaje proporciona la base empírica necesaria. Para poder obtener conclusiones metafísicas de una teoría de la verdad de la manera que yo propongo, la aproximación al lenguaje debe ser holística. Sin embargo, el propio Quine pasa por alto esta significación metafísica tan directa del holismo, y ello por una cantidad de razones. Primero, Quine no ha considerado esencial la teoría de la verdad ni como una clave de la ontología de un lenguaje, ni como un test de la forma lógica. Segundo, él, como Frege, ve un lenguaje satisfactoriamente regimentado como un mejoramiento del lenguaje natural más que como una parte de una teoría acerca de él. En un aspecto importante, Quine parece incluso ir más allá de Frege, pues donde Frege piensa que su notación conduce a un mejor lenguaje, Quine piensa que conduce también a una ciencia mejor. Por consecuencia, Quine liga su metafísica a su notación canónica más que al lenguaje natural; tal como él lo dice: "La búsqueda de un modelo general de notación canónica más simple y más claro no puede distinguirse de una búsqueda de categorías últimas, una iluminación de los trazos más generales de la realidad."[1]

[1] W. V. Quine, *Word and Object*, 161.

Los lenguajes formales por los cuales yo me inclino —lenguajes de primer orden con lógica usual— son los mismos que prefiere Quine. Pero nuestras razones para esta elección divergen en alguna medida. Estos lenguajes gustan a Quine porque su lógica es simple, y las porciones científicamente respetables del lenguaje natural pueden ser traducidas a ellos; y en esto yo estoy de acuerdo. Pero dado que yo no estoy interesado en mejorar el lenguaje natural, sino en comprenderlo, entiendo los lenguajes formales o las notaciones canónicas como recursos para explorar la estructura del lenguaje natural. Sabemos cómo dar una teoría de la verdad para el lenguaje formal; por lo tanto, si supiéramos también cómo transformar sistemáticamente las oraciones de un lenguaje natural en oraciones del lenguaje formal, tendríamos una teoría de la verdad para el lenguaje natural. Desde este punto de vista, los lenguajes formales usuales son recursos intermedios para ayudarnos a tratar a los lenguajes naturales como lenguajes formales más complejos.

El trabajo de Tarski en torno de las definiciones de la verdad para los lenguajes formalizados sirve como inspiración para la clase de teoría de la verdad que se quiere para los lenguajes naturales.[2] El método opera enumerando las propiedades semánticas de los ítems de un vocabulario finito, y en base a esto caracteriza recursivamente la verdad para cada una de las infinitas oraciones. La verdad se alcanza desde ese punto de partida merced a la intervención de concepto sutil y poderoso (la satisfacción) que relaciona tanto las oraciones como las expresiones no oracionales con objetos del mundo. Un rasgo importante de la aproximación de Tarski es que sólo se acepta una caracterización de un predicado de verdad "x es verdadera en L" si ella implica, para cada oración del lenguaje L, un teorema de la forma "x es verdadera en L si y sólo si..." donde "x" se reemplaza por una descripción de la oración y los puntos se reemplazan por una traducción de la oración al lenguaje de la teoría.

Resulta evidente que estos teoremas, que podemos llamar oraciones T, requieren un predicado que valga sólo para las oraciones verdaderas de L. También salta a la vista, a partir del hecho de que las condiciones de verdad para una oración traducen esa oración (por ejemplo, lo que aparece a la derecha de "si y sólo si" en una oración T traduce la oración descrita a la izquierda), que la teoría muestra cómo caracterizar la verdad para cualquier oración dada sin apelar a recursos conceptuales no disponibles en esa oración.

Estas afirmaciones son correctas sólo a grandes rasgos. Una teoría de la verdad para un lenguaje natural debe relativizar la verdad de una oración a las circunstancias de emisión, y cuando esto se ha hecho las condiciones de verdad dadas por una oración T ya no traducirán la oración descrita, ni será posible evitar el empleo de conceptos que son, quizá, semánticos, para dar las condiciones de verdad de oraciones con elementos deícticos. Más importante aun, la noción de traducción, que puede precisarse para los lenguajes artificiales a los cuales pueden im-

[2] A. Tarski, "The Concept of Truth in Formalized Languages".

ponerse interpretaciones por decreto, no tiene una aplicación precisa y ni siquiera clara para los lenguajes naturales.

Por estas y otras razones, es importante destacar que una teoría de la verdad para un lenguaje natural (tal como yo la concibo) difiere ampliamente de las definiciones de la verdad de Tarski tanto en su finalidad como en su interés. La precisión de su aplicación se pierde, y con ella la mayor parte de lo que concierne a matemáticos y lógicos: consecuencias para consistencia, por ejemplo. Tarski podía considerar que la traducción estaba sintácticamente especificada, y seguir adelante para definir la verdad. Pero en la aplicación a un lenguaje natural tiene más sentido suponer una comprensión parcial de la verdad y usar la teoría para echar luz sobre el significado, la interpretación y la traducción.[3] La satisfacción de la Convención T de Tarski permanece como un desiderátum de una teoría, pero ya no se halla disponible como test formal.

Una teoría de la verdad logra revelar la estructura de un lenguaje natural. Al tratar a cada oración como compuesta de maneras explicables a partir de un número finito de palabras veritativas relevantes, ella articula esta estructura. Cuando estudiamos términos y oraciones directamente, no a la luz de una teoría abarcadora, debemos llevar metafísica al lenguaje; asignamos roles a las palabras y las oraciones de acuerdo con las categorías que postulamos independientemente en base a fundamentos epistemológicos o metafísicos. Al operar de esta manera, los filósofos ponderan cuestiones tales como si debe haber entidades, tal vez universales, que correspondan a predicados, o entidades no existentes que correspondan a nombres o descripciones no denotativos; o discuten si las oraciones corresponden o no corresponden a hechos o proposiciones.

Una luz diferente ilumina estos asuntos cuando buscamos una teoría abarcadora de la verdad, pues una teoría tal expresa sus propios requerimientos insoslayables.

III

Consideremos ahora algunas aplicaciones. Ya notamos que el requisito de que las condiciones de verdad de una oración sean dadas usando solamente los recursos conceptuales de esa oración no aclara del todo dónde es pasible de ser cumplido, ni parece ser aplicable para todos los casos. Los casos que invitan a la excepción son las oraciones que involucran demostrativos, y aquí la cura de la dificultad es relativamente simple.[4] Al margen de estos casos, a mi modo de ver, el requisito tiene, a pesar de su oscuridad, importantes implicaciones.

Supónganse que admitiéramos una regla así como parte de una teoría de la verdad: "Una oración consistente en un término singular se-

[3] Véanse los Ensayos 9 y 10.
[4] Véase S. Weisntein, "Truth and Demonstratives".

guido de un predicado monádico es verdadera si y sólo si el objeto nombrado por el término singular pertenece a la clase determinada por el predicado."[5] Esta regla constituye una afrenta para el requisito, pues si se la admitiera, la oración T para "Sócrates es sabio" sería " 'Sócrates es sabio' es verdadero si y sólo si el objeto nombrado por 'Sócrates' pertenece a la clase determinada por el predicado 'es sabio' ", y aquí el enunciado de las condiciones de verdad involucra dos conceptos semánticos (nombrar y determinar una clase) que plausiblemente no se encontrarán entre los recursos conceptuales de "Sócrates es sabio".

Sería fácil pasar de la tendenciosa oración T recién mencionada a la no comprometida y admisible " 'Sócrates es sabio' es verdadero si y sólo si Sócrates es sabio" si la teoría contuviera también como postulados enunciados de que el objeto nombrado por "Sócrates" es Sócrates y de que x pertenece a la clase determinada por el predicado "es sabio" si y sólo si x es sabio. Si se dispone de la cantidad suficiente de tales postulados como para cuidar de todos los nombres propios y predicados primitivos, los resultados son claros. Primero, habría oraciones T libres de términos semánticos indeseables a disposición de todas las oraciones involucradas; y los términos semánticos extra serían innecesarios. Pues tendría que haber un postulado para cada nombre y predicado, y esto sólo podría ocurrir si la lista de nombres y predicados primitivos fuera finita. Pero si la lista fuera finita, sólo habría un número finito de oraciones consistentes en un nombre y un predicado monádico, y nada se interpondría en el camino para dar directamente las condiciones de verdad de todas esas oraciones; las propias oraciones T podrían servir de axiomas.

El ejemplo ilustra la forma en que el mantener finito el vocabulario puede dar lugar a la eliminación de conceptos semánticos; muestra también cómo la exigencia de una teoría satisfactoria tiene consecuencias ontológicas. Aquí, la apelación a entidades que correspondan a predicados desaparece cuando se hace que la teoría produzca oraciones T sin un bagaje semántico excesivo. Ciertamente en el caso de marras la teoría no necesita en absoluto expresiones y objetos de una correspondencia explícita, y por lo tanto no se incluye ontología alguna; pero esto se debe a que la provisión de oraciones cuyas condiciones de verdad deben darse es finita.

No es que una infinitud de oraciones requiera necesariamente ontología. Dada la provisión finita de oraciones con predicados inestructurales que hemos estado imaginando, es fácil proseguir hacia el infinito mediante la adición de uno o más recursos iterables para la construcción de oraciones a partir de oraciones, como la negación, la conjunción o la alternación. Si no se requirió ontología para dar las condiciones de verdad de las oraciones más simples, estos recursos tampoco tendrán mayores necesidades.

Sin embargo, por lo general la estructura semánticamente relevan-

[5] Véase R. Carnap, *Meaning and Necessity*, 5.

te tiende a demandar ontología. Considérese, por ejemplo, la idea de que las citas deben ser tratadas como átomos semánticos, a la par de los nombres propios en cuanto a su carencia de estructura significativa. Dice Tarski que esta forma de ver la cita "parece ser la más natural y en completa concordancia con la forma acostumbrada de usar las comillas".[6] Este autor ofrece un argumento modelo para mostrar que las comillas no pueden ser tratadas como una expresión funcional ordinaria puesto que una cita no nombra una entidad que sea una función de algo nombrado por lo que las comillas encierran. Sin duda Tarski tiene razón acerca de esto, pero la moraleja de la lección no puede ser que las citas son similares a los nombres propios; en cualquier caso, no lo será si es que puede darse una teoría de la verdad al estilo de Tarski para un lenguaje que contenga citas. Pues está claro que hay infinitas citas.

Puede extraerse una idea para una posible solución a partir de la afirmación de Quine de que las citas pueden reemplazarse por el deletreo (Tarski dice casi lo mismo). El deletreo sí tiene estructura. Es una manera de dar una descripción semánticamente articulada de una expresión mediante el uso de un número finito de expresiones: el signo de concatenación, con paréntesis asociados, y los nombre (propios) de las letras. Siguiendo esta línea, deberíamos pensar que una cita como " 'gato' " tiene una forma que está dada más claramente por " 'g'⌒'a'⌒'t'⌒'o' ", o, mejor aun, por "ge⌒a⌒te⌒o". Esta idea funciona, al menos hasta cierto punto. Pero veamos las consecuencias. Ya no consideramos inestructurada a la cita " 'gato' "; más bien la tratamos como una abreviación de algún tipo para una descripción compleja. Mas no una abreviación arbitraria a ser especificada para el caso en estudio, sino como un *estilo* de abreviación que puede ser expandido mecánicamente a una descripción que muestre la estructura más claramente. Por cierto, hablar de abreviación puede resultar engañoso; bien podemos decir que esta teoría trata a las citas como descripciones complejas.

Otra consecuencia es que al dar estructura a las citas tuvimos que reconocer en ellas "palabras" repetibles e independientes: nombres de las letras individuales y el signo de concatenación. Estas "palabras" son, desde luego, finitas en número —tal era el requerimiento— pero también revelan un hecho ontológico que no era aparente cuando se consideraba a las citas como nombres inestructurados: un compromiso con las letras. Tenemos una teoría manejable cuando explicamos que las moléculas están hechas de átomos de un número finito de clases; pero también tenemos los átomos.

Tenemos otro ejemplo aun más impactante de la forma en que la postulación de estructura necesaria en el lenguaje puede conllevar la aportación de ontología, proporcionado por la semántica de Frege para los contextos oblicuos creados por oraciones acerca de actitudes proposicionales. Para Frege, una oración como "Daniel cree que hay un león en

[6] A. Tarski, "The Concept of Truth in Formalized Languages", 160. Si se desea más material sobre la cita véase el Ensayo 6.

la guarida" está dominada por el predicado diádico "cree" cuyo primer espacio está ocupado por el término singular "Daniel" y cuyo segundo espacio está ocupado por un término singular que nombra una proporción o "sentido". Para seguir esta idea no sólo es menester que tratemos a las oraciones como términos singulares, sino que también hallemos entidades para que nombren. Y aun hay más. Pues está claro que un número infinito de oraciones puede ocupar el puesto que sigue a "Daniel cree que..." Es así como para proveer una definición de la verdad, debemos descubrir estructura semántica en estos términos singulares: debe mostrarse cómo se los puede tratar como descripciones de proposiciones. Para evitar los absurdos que sobrevendrían si los términos singulares de una oración tuvieran su referencia usual, Frege considera en cambio que refieren entidades intensionales. Debe operarse cambios análogos sobre las características semánticas de predicados, cuantificadores y conectivos oracionales. Hasta el momento, una teoría de la verdad del tipo que hemos estado buscando puede manejar la situación, pero sólo si considera que cada palabra del lenguaje es ambigua, con una interpretación en contextos ordinarios y otra cuando sigue a "cree *that*" y verbos similares. Lo cual vendría a significar que, como ventaja de esta teoría, una palabra debe tratarse como dos. Frege notó esto, y registró la ambigüedad como un punto en contra del lenguaje natural; Church, en los lenguajes artificiales de *"A Formulation of the Logic of Sense and Denotation"*, eliminó la ambigüedad introduciendo expresiones distintas, que diferían en el subíndice.[7]

Frege sugirió que con cada agregado de un verbo de actitud proposicional delante de una expresión referente esa expresión pasa a referir una entidad de nivel semántico mayor. De esta manera, cada palabra y oración es infinitamente ambigua en varias formas; para la teoría de Church habrá un vocabulario básico infinito. Ni en uno ni en otro caso es posible proporcionar una teoría de la verdad del tipo que queremos.

Frege fue claro respecto de la necesidad —si es que queremos contar con una teoría sistemática— de ver el valor de verdad de cada oración como una función de los roles semánticos de sus partes o aspectos, mucho más claro que cualquiera que lo haya precedido, y más claro que la mayoría de quienes lo siguieron. Lo que Frege no advirtió, según surge de este último ejemplo, fueron las restricciones adicionales, en particular respecto de un vocabulario finito, que emanan del requisito de una amplia teoría de la verdad. Frege llevó la semántica al punto donde el requisito se tornó inteligible e incluso, tal vez, satisfacible; pero no se le ocurrió formularlo.

Observemos más en detalle la operación que nos permite sacar a la luz la estructura latente mediante la caracterización de un predicado de verdad. Los primeros pasos pueden ilustrarse con una oración tan simple como "Juan y Marina subieron la colina": ¿bajo qué condiciones es esta oración verdadera? El desafío descansa en la presencia dentro de la

[7] A. Church, "A Formulation of the Logic of Sense and Denotation".

oración de un recurso iterativo: la conjunción. Resulta claro que podemos seguir agregando frases como "y Susana" después de la palabra "Marina" *ad libitum*. Por lo tanto cualquier enunciado de las condiciones de verdad para esta oración debe tener en cuenta la infinitud de oraciones, generadas por el mismo recurso, que están esperando su tratamiento. Lo que hace falta es una cláusula recursiva en la teoría de la verdad que pueda aplicarse cuantas veces sea necesaria. La treta, como todos sabemos, consiste en definir la verdad para un repertorio básico y finito de las oraciones más simples, tales como "Juan subió la colina" y "Marina subió la colina", y luego hacer que las condiciones de verdad para "Juan y Marina subieron la colina" dependan de las condiciones de verdad de las dos oraciones simples. Así obtenemos:

> "Juan y Marina subieron la colina" es verdadero si y sólo si Juan subió la colina y Marina subió la colina.

como una consecuencia de una teoría de la verdad. A la izquierda, una oración del lenguaje vernáculo se describe con su estructura transparente o no; a la derecha del "si y sólo si" tenemos una oración del mismo lenguaje vernáculo, aunque se ha elegido de éste una parte capaz de hacer explícita, merced a aplicaciones repetidas de los mismos recursos simples, la estructura semántica subyacente. Si una teoría de la verdad produce una oración así purificada para cada oración del lenguaje, la porción del lenguaje total que se usa a la derecha puede considerarse una notación canónica. Ciertamente, si se sustituyen algunas palabras por símbolos y se evidencian los agrupamientos por medio de paréntesis o algún recurso equivalente, la parte del lenguaje empleada en establecer las condiciones de verdad para todas las oraciones puede volverse indistinguible de lo que a menudo se llama un lenguaje formalizado o artificial. Sin embargo, sería un error suponer que es esencial hallar una subdivisión canónica de este tipo en el lenguaje. Puesto que "y" puede escribirse entre oraciones en el castellano, tomamos la vía fácil de transformar "Juan y Marina subieron la colina" en "Juan subió la colina y Marina subió la colina" y luego dar las condiciones de verdad de esta última de acuerdo con una regla que dice que una conjunción de oraciones es verdadera si y sólo si cada conyunto lo es. Pero supongamos que "y" nunca haya estado presente entre oraciones; su papel como conectivo oracional aun se vería reconocido por una regla que dice que una oración compuesta por un sujeto conjuntivo ("Juan y Marina") y un predicado ("subieron la colina") es verdadera si y sólo si la oración compuesta por el primer sujeto de la conjunción y el predicado, y la oración compuesta por el segundo sujeto de la conjunción y el predicado, son verdaderas. La regla que se requiere para cumplir la tarea de la simple regla original es menos perspicua y necesita ser suplementada con otras. Pero la cuestión no varía: la notación canónica es una comodidad de la que, llegado el caso, podemos prescindir. Es bueno, pero no necesario, llevar forma lógica a la superficie.

De manera similar, el tratamiento de la negación se vería facilita-

do en gran medida si pudiéramos transformar de un modo plausible todas las oraciones que contienen negación en oraciones con el mismo valor de verdad y en las cuales la frase de negación siempre gobierne una oración (como con "no es el caso que"). Pero si esto no fuera posible, la negación aún sería un conectivo oracional si la condición de verdad de una oración como "El carbón no es blanco" se diera haciendo referencia a la condición de verdad de "El carbón es blanco". ("El carbón no es blanco" es verdadero si y sólo si "El carbón es blanco" no es verdadero.)

El tema de la ontología se ve forzado a salir a la luz sólo donde la teoría encuentra estructura cuantificacional, y allí es donde la teoría explica mejor el patrón de dependencias de la verdad, relacionando sistemáticamente expresiones con objetos. Causa sorpresa cuán firmemente la exigencia de teoría llama a sosiego a una antigua fuente de aporía: la cuestión de cómo demostrar la asimetría de sujeto y predicado. Hasta tanto nuestra atención se centre en oraciones únicas y simples, podemos preguntarnos por qué una explicación de la verdad tendría que involucrar a los predicados en la ontología en menor grado que los términos singulares. La clase de los objetos sabios (o la propiedad de la sabiduría) ofrece por sí misma cuanto podría corresponder al predicado "sabio" en "Sócrates es sabio" de una manera muy parecida a como Sócrates corresponde a "Sócrates". Como señaláramos antes, ningún número finito de tales oraciones requiere que una teoría de la verdad ponga ontología en escena. Sin embargo, cuando llegamos a la cuantificación mixta y a predicados con algún grado de complejidad, la escena cambia. Con una estructura cuantificacional compleja, la teoría hará coincidir expresiones con objetos. Pero, siempre y cuando se suponga que la lógica subyacente sea de primer orden, no hay necesidad de introducir entidades para corresponder a los predicados. Desde ya, el reconocimiento de este hecho no pondrá fin a la cuestión de si existen o no cosas tales como los universales y las clases. Pero sí demuestra que hay una diferencia entre un término singular y un predicado; para grandes extensiones del lenguaje, de todas maneras, las variables, los cuantificadores y los términos singulares deben interpretarse como referenciales en cuanto a su función; esto no es así para los predicados.

No siempre resulta evidente cuál es la estructura cuantificacional de una oración del lenguaje natural; lo que parecen ser términos singulares a veces se funden en algo de implicancias menos ónticas cuando se estudian sus relaciones lógicas con otras oraciones, mientras que los requerimientos de teoría pueden sugerir que una oración desempeña un papel que puede explicarse solamente si se la trata como si tuviera una estructura cuantificacional no aparente en su superficie. He aquí un ejemplo sencillo.

¿Cuál es la ontología de una oración como

"Juan se cayó pero luego Juan se rompió la coronilla"?

Juan y su coronilla parecen ser los únicos candidatos a entidades que deben existir para que esta oración sea verdadera. Y si en lugar de

"pero luego" tuviéramos "y", esta respuesta nos podría satisfacer por la razón ya explorada: es decir, que podemos enunciar, de un modo que funcionará para innumerables casos similares, las condiciones de verdad de toda la oración "Juan se cayó *y* Juan se rompió la coronilla", precisamente en base a la verdad de las oraciones componentes, y que podemos esperar dar las condiciones de verdad para las componentes sin más ontología que Juan y su coronilla. Pero "Juan se cayó pero luego Juan se rompió la coronilla" no se somete a este tratamiento, pues "pero luego" no puede considerarse como un conectivo semántico veritativo funcional: para entender esto, reflexiónese que para que la oración sea verdadera, ambas oraciones componentes deben ser verdaderas, pero esto no es suficiente para su verdad, puesto que intercambiando las componentes se la hará falsa.

Frege nos indicó cómo habérnoslas con un caso así: podemos formular las condiciones de verdad para la oración "Juan se cayó pero luego Juan se rompió la coronilla" como sigue: ella es verdadera si y sólo si existe un tiempo t y existe un tiempo t' tal que Juan se cayó en t, Juan se rompió la coronilla en t', y t está antes que t'. Aparentemente nos sometemos a la existencia de tiempos si aceptamos como verdadera una oración así. Y pensando en el carácter holístico de una definición de la verdad, el descubrimiento de ontología oculta en oraciones que contienen "luego" debe transferirse a otras oraciones: de esta manera, "Juan se cayó" es verdadero si y sólo si existe un tiempo t tal que Juan se cayó en t.

Pasemos ahora a un ejemplo más perturbador. Consideremos primero "La caída de Juan causó la rotura de su coronilla". Aquí es natural tomar "La caída de Juan" y "la rotura de su coronilla" como términos singulares que describen sucesos, y "causó" como un predicado diádico o relacional. Pero entonces, ¿cuál es la relación semántica entre términos generales como "caída" en "La caída de Juan" y verbos como "se cayó" en "Juan se cayó"? En cuanto a eso, ¿cómo difiere "La caída de Juan causó la rotura de su coronilla", en sus condiciones de verdad, de "Juan se cayó, lo que causó que se diera el caso que Juan se rompiera la coronilla", donde la frase "lo que causó que se diera el caso que" es, por lo visto, un conectivo oracional?

La teoría correcta de "causó", como ya ha sido discutido más extensamente en otra parte, es paralela a "luego" o a la teoría de Frege para "antes".[8] Sugiero que "Juan se cayó, lo que causó la rotura de su coronilla" es verdadero si y sólo si existen sucesos e y f tal que e es la caída que tuvo Juan, f es la rotura sufrida en su coronilla, y e causó f. De acuerdo con esta propuesta, el predicado "es una caída", consistente con los sucesos, se torna primario, y los contextos que contienen el verbo son derivados. De esta manera "Juan se cayó" es verdadero si y sólo si hay una caída tal que aconteció a Juan, "Juan dio un paseo" es verdadero si y sólo si hay un paseo que dio Juan, y así por el estilo. Según este análi-

[8] Véase el Ensayo 7 en *Essays on Actions and Events*.

sis, una frase nominal como "La caída de Juan" se vuelve una descripción genuina, y lo que ella describe es esa caída que tuvo Juan.

Una consideración que puede ayudar a reconciliarnos con una ontología de sucesos particulares es que podemos luego prescindir de la ontología abstracta de tiempos que acabamos de aceptar tentativamente, pues los sucesos son tan admisiblemente los *relata* de la relación anterior como lo son los tiempos. Otra consideración es que reconociendo nuestro compromiso con una ontología de sucesos podemos hallar el camino hacia una semántica visible de adverbios y de modificación adverbial. Sin sucesos, aparece el problema de explicar las relaciones lógicas entre oraciones como "Jones se cortó la mejilla mientras se afeitaba con una navaja en el baño este sábado", y "Jones se cortó la mejilla en el baño", y "Jones se cortó la mejilla". Parece que está operando algún recurso iterativo; pero ¿qué puede ser ese recurso, desde el punto de vista semántico? Los libros de lógica no lo dicen: al analizar estas oraciones afirman que ellas requieren relaciones con números de espacios variables dependiendo del número de modificaciones adverbiales, pero esto conduce a la conclusión inaceptable de que hay un vocabulario básico infinito, y no logra explicar las obvias inferencias. Si interpretamos que estas oraciones tratan de eventos, podemos resolver los problemas. Luego podemos decir que "Jones se cortó la mejilla en el baño este sábado" es verdadero si y sólo si existe un evento que consiste en un corte de su mejilla por parte de Jones, *y* ese evento tuvo lugar en el baño, *y* tuvo lugar este sábado. El recurso iterativo es ahora obvio: consiste en la colaboración familiar entre conjunción y cuantificación que nos permite tratar con "alguien se cayó y se rompió la coronilla".

Este recurso funciona, pero como hemos visto, hace falta una ontología para hacerlo funcionar: una ontología que incluya gente para "Alguien se cayó y se rompió la coronilla", una ontología de sucesos (además) para "Jones se cortó la mejilla en el baño este sábado". Resulta ligeramente irónico que en la filosofía reciente se haya vuelto una maniobra popular el intento de *evitar* problemas ontológicos tratando determinadas frases como si fueran adverbiales. Una de estas sugerencias dice que podemos abjurar de los *sense-data* si convertimos una oración como "La montaña aparece azul a los ojos de Smith" en "La montaña aparece azuladamente a los ojos de Smith". Otra idea similar dice que podemos pasarla sin una ontología de objetos intensionales si pensamos que las oraciones acerca de actitudes proposicionales son esencialmente adverbiales: "Galileo dijo que la Tierra se mueve" se convertiría entonces en "Galileo habló en una forma que-la-Tierra-se-mueve". A mi entender, las posibilidades de que estas cláusulas adverbiales puedan recibir un análisis semántico sistemático sin complicaciones ontológicas son mínimas.

Hay otra vía, muy distinta, por la cual una teoría de la verdad puede tener repercusiones metafísicas. Al ajustarse a la presencia de demostrativos, y de elementos demostrativos como el tiempo verbal, en un lenguaje natural, una teoría de la verdad debe tratar a la verdad como un atributo de las emisiones que depende (quizás entre otras cosas) de

la oración emitida, del hablante y del tiempo. Alternativamente, sería posible tratar a la verdad como una relación entre hablantes, oraciones y tiempos. Luego una emisión de "Tengo un metro y medio de altura" es verdadera si se la dice en ciertos tiempos en las vidas de la mayor parte de la gente, y es verdadera si se la dice en cualquier tiempo durante un período considerable de las vidas de unos pocos. "Se te ven los calzoncillos" puede ser verdadera cuando es emitida por un hablante en un momento en que está mirando hacia el oeste, aunque podría no haber sido verdadera si hubiera estado mirando hacia el norte; y "Hilary ascendió al Everest" fue durante largo tiempo falsa, y es hoy verdadera para siempre. Las oraciones sin elementos demostrativos no pueden hacer el trabajo de las oraciones con elementos demostrativos, pero si es que queremos tener una teoría de la verdad debemos ser capaces de enunciar, sin el uso de demostrativos, una regla que explique bajo qué condiciones las oraciones con demostrativos son verdaderas. Esas reglas darán la condición de verdad de oraciones como "Hilary ascendió al Everest" solamente cuantificando sobre emisiones, hablantes y tiempos o, tal vez, sucesos.

Si para dar una teoría de la verdad debe apelarse explícitamente a los hablantes y sus circunstancias, entonces, en base a la suposición de que las características generales del lenguaje reflejan características objetivas del mundo, debemos concluir que una metafísica inteligente asignará un lugar preponderante a la idea de gente (=hablantes) con una posición en el espacio y el tiempo públicos.

Debe quedar claro que "el método de la verdad" en metafísica no elimina el recurso a argumentos y decisiones más comunes y a menudo esencialmente no lingüísticos. Lo que sea posible hacer dentro de una teoría de la verdad, por ejemplo, depende en gran medida de los recursos lógicos que la teoría misma despliega, y la teoría no puede decidir esto por nosotros. Como hemos visto, el método tampoco sugiere qué verdades, más allá de las que descuenta como lógicas, debemos aceptar como una condición para la comprensión mutua. Lo que hace una teoría de la verdad es describir el patrón que la verdad debe hacer efectivo entre las oraciones, sin decirnos cuándo ese patrón fracasa. Así, por ejemplo, yo argumento que un número muy grande de nuestras afirmaciones ordinarias acerca del mundo no puede ser verdadero a menos que haya sucesos. Pero una teoría de la verdad, aun si tomara la forma que yo propongo, no especificaría qué sucesos existen, ni siquiera que existe alguno. Sin embargo, si yo estoy en lo cierto respecto de la forma lógica de las oraciones que conciernen al cambio, luego a menos que haya sucesos no hay oraciones verdaderas de tipos muy comunes acerca del cambio. Y si no hay oraciones verdaderas acerca del cambio, no hay oraciones verdaderas acerca de objetos que cambian. Un metafísico que desee suponer que ninguna oración como "El Vesubio entró en erupción en marzo de 1944" o "César cruzó el Rubicón" es verdadera no se verá forzado por una teoría de la verdad a admitir la existencia de sucesos o hasta quizá de personas o de montañas. Pero si acepta que muchas de esas oraciones son verdaderas (sean ellas cuales fueren), luego resulta obvio que

debe aceptar la existencia de personas y de volcanes; y, si estoy en lo cierto, debe aceptar la existencia de sucesos como las erupciones y los cruces.

El mérito del método de la verdad no consiste en que resuelva tales cuestiones de una vez y para siempre, ni tampoco en que las resuelva sin más reflexión metafísica. Pero el método sirve efectivamente para pulir nuestro sentido de las alternativas viables, y proporciona una amplia idea de las consecuencias de una decisión. La metafísica tiene la generalidad por meta; el método de la verdad expresa esa exigencia mediante el requerimiento de una teoría que alcance a todos los fundamentos. De este modo los problemas de la metafísica, mientras no se los haya resuelto o reemplazado, pasar a ser vistos como los problemas de toda construcción de buenas teorías. Queremos una teoría que sea simple y clara, con un aparato lógico que se comprenda y se justifique, y que se ocupe de los hechos relacionados con el funcionamiento de nuestro lenguaje. Que sean esos hechos, es algo que de alguna manera seguirá en disputa, como seguramente sucederá respecto de las diversas discusiones de lo que es simplicidad y claridad. Estas preguntas serán, no lo dudo, las viejas preguntas de la metafísica como un vestido nuevo. Pero el nuevo vestido resulta en muchos aspectos muy atractivo.

15

Realidad sin referencia

Resulta difícil de comprender cómo una teoría del significado que no dilucide y ofrezca un papel central al concepto de referencia puede tener esperanzas de éxito. Por otra parte, hay razones de peso para suponer que la referencia no puede explicarse o analizarse en términos más primitivos o conductales. Permítaseme describir el dilema de manera más completa, y luego decir cómo pienso que una teoría de la verdad al estilo de Tarski puede ayudar a resolverlo.

"Teoría del significado" no es un término técnico, sino una indicación en dirección a una familia de problemas. Entre estos problemas es fundamental la tarea de explicar el lenguaje y la comunicación apelando a conceptos más simples o en todo caso diferentes. Es natural creer que esto es posible porque, claramente, los fenómenos lingüísticos hacen su aparición después de los fenómenos no lingüísticos. Propongo que una teoría pueda llamarse teoría del significado para un lenguaje natural *L* si es tal que (a) el conocimiento de la teoría es suficiente para comprender las emisiones de los hablantes de *L* y (b) la teoría puede recibir aplicación empírica apelando a la evidencia descrita sin el uso de conceptos lingüísticos, o al menos sin el uso de conceptos lingüísticos específicos de las oraciones y palabras de *L*. La primera condición indica la naturaleza de la cuestión; la segunda requiere que no se la dé por supuesta.

Cuando digo teoría de la verdad, quiero significar una teoría que satisfaga algo como la Convención T de Tarski: es una teoría que mediante la caracterización recursiva de un predicado de verdad (digamos, "es verdadero en *L*") implica, para cada oración *s* de *L*, una oración metalingüística obtenida a partir de la forma "*s* es verdadera en *L* si y sólo si p" cuando "*s*" se reemplaza por una descripción canónica de una oración de *L* y "p" por una oración del metalenguaje que dé las condiciones de verdad de la oración descrita. La teoría debe relativizarse a un tiempo y a un hablante (por lo menos) para poder manejar las expresiones deícticas. No obstante llamaré *absolutas* a estas teorías para distinguirlas de las teorías que (también) relativizan la verdad a una interpretación, un modelo, un mundo posible o un dominio. En una teoría del tipo que estoy describiendo, el predicado de verdad no está definido, pero se lo debe considerar una expresión primitiva.

Podemos entender que la referencia es una relación entre nombres propios y lo que ellos nombran, entre términos singulares complejos y lo

que ellos denotan, entre predicados y las entidades para las cuales ellos son verdaderos. Los demostrativos no entrarán en la discusión, pero, desde luego, sus referencias tendrían que relativizarse a un hablante y un tiempo (al menos).

Volvamos al dilema. Es aquí donde parece que no podemos pasárnosla sin el concepto de referencia. Además de cualquier otro punto que pueda abarcar, una teoría del significado debe incluir una consideración de la verdad: un enunciado de las condiciones bajo las cuales una oración arbitraria del lenguaje es verdadera. Por razones bien conocidas, una tal teoría no puede comenzar explicando la verdad para un número finito de oraciones simples y luego asignar verdad al resto de las oraciones en base a las simples. Es necesario, y en cualquier caso sería deseable, analizar las oraciones en sus elementos constituyentes —predicados, nombres, conectivos, cuantificadores, functores— y mostrar la forma en que el valor de verdad de cada oración deriva de rasgos de los elementos y de la composición de los elementos en la oración. Luego, la verdad depende claramente de los rasgos semánticos de los elementos: y donde los elementos son nombres o predicados, ¿qué rasgos pueden ser relevantes sino la referencia? Explicar las condiciones de verdad de una oración como "Sócrates vuela" debe equivaler a decir que es verdadera si y sólo si el objeto referido por "Sócrates" es uno de los objetos referidos por el predicado "vuela".

Una teoría de la verdad del tipo que acabo de mencionar muestra efectivamente cómo las condiciones de verdad de cada oración son una función de los rasgos semánticos de los ítem de un vocabulario básico finito. Pero se dice a menudo que una teoría así no explica los rasgos semánticos del vocabulario básico. En una teoría de la verdad hallamos aquellas cláusulas recursivas familiares que especifican, por ejemplo, que una conjunción es verdadera si y sólo si cada conyunto es verdadero, que una disyunción es verdadera si y sólo si al menos uno de los disyuntos es verdadero, y así por el estilo. (De hecho la teoría debe explicar cómo trabajan los conectivos tanto en oraciones cerradas como en oraciones abiertas, y por lo tanto la recursión se aplicará a la relación de *satisfacción* antes que directamente a la verdad.)

Sabemos que la teoría debe implicar una oración T aun para los casos más simples, como por ejemplo:

(T) "Sócrates vuela" es verdadera si y sólo si Sócrates vuela.

Si nada se dice de los constituyentes, ¿cómo se las arregla la teoría con casos así? Bueno, una forma podría ser ésta: el vocabulario básico debe ser finito. En particular, luego, sólo puede haber un número finito de predicados simples y un número finito de nombres propios (términos singulares inestructurados, sin contar las variables). Entonces es posible hacer un listado de oraciones consistentes en un nombre propio y un predicado básico. Se sigue que una teoría puede implicar cada oración del tipo (T) gracias al hecho de tener cada oración así como axioma. Resulta claro que este método evita (hasta ahora) toda

apelación al concepto de referencia... ni logra echar luz alguna sobre el mismo.

Los predicados se presentan en cualquier grado de complejidad, puesto que se los puede construir a partir de conectivos y de variables; y los términos singulares constantes pueden ser complejos. Luego el método que estábamos explorando recién fracasará en general en sus funciones. En el caso de los predicados, el método de Tarski, como sabemos, involucra el recurso al concepto de satisfacción, una relación entre predicados y n-tuplos de entidades para las cuales los predicados son verdaderos (en realidad, se trata de secuencias de los mismos). La satisfacción es obviamente muy parecida a lo que es la referencia para los predicados: en efecto, podríamos definir la referencia de un predicado como la clase de aquellas entidades que lo satisfacen. El problema está en que una teoría absoluta de la verdad no ilumina realmente la relación de satisfacción. Cuando la teoría viene a caracterizar la satisfacción para el predicado "x vuela", por ejemplo, nos dice meramente que una entidad satisface "x vuela" si y sólo si esa entidad vuela. Si solicitamos una mayor explicación o análisis de la relación, nos veremos decepcionados.

El hecho de que una definición absoluta de la verdad no logre producir un análisis del concepto de referencia puede verse a partir de lo siguiente: si uno imagina un nuevo predicado sumado al lenguaje —o un lenguaje exactamente igual al viejo, excepto por contener un único predicado más— la consideración de la verdad y de la satisfacción ya dadas no sugieren *cómo pasar al nuevo caso.* (Esta afirmación no es aplicable a las cláusulas recursivas: ellas dicen *en general* cuándo es verdadera una conjunción, no importa cuáles los conyuntos.)

El hecho de que la satisfacción, a la que hemos estado considerando caracterizada recursivamente, pueda recibir una definición explícita (mediante la técnica de Frege-Dedekind) no debe llevarnos a pensar que se ha captado un concepto general. Y esto porque la definición (al igual que la recursión que la sirve) limitará explícitamente la aplicación de la satisfacción a una lista finita y fija de predicados (y compuestos de ellos). Así, si una teoría (o definición) de la satisfacción se aplica a un lenguaje dado y luego se agrega un nuevo predicado, digamos, "x vuela", sucederá que "x vuela" no se verá satisfecho por un objeto que vuela, ni por ninguna otra cosa.

Afirmaciones análogas valen para los términos singulares constantes. Ciertamente, si hay términos singulares complejos será necesario caracterizar una relación como la referencia, usando cláusulas recursivas como: "el padre de" concatenado con un nombre α refiere al padre de lo que α refiere. Pero para los nombres propios básicos, simplemente habrá otra vez una lista. No se analizará qué es para un nombre propio referir a un objeto.

El punto que acabo de desenterrar —la existencia de un claro sentido en el cual una teoría absoluta de la verdad no echa luz sobre los rasgos semánticos del vocabularios básico de predicados y nombres— tiene un aire familiar. Esta queja se ha sumado muchas veces a otra, se-

gún la cual una teoría de la verdad al estilo de Tarski no logra penetrar en el concepto de verdad. No nos hará ningún daño aceptar la idea de que en una teoría de la verdad, la expresión "es verdadero" (o cualquiera que tome su lugar) se comprende de manera independiente. Las razones por cuales la Convención T resulta aceptable como criterio de teorías son (1) que las oraciones T son claramente verdaderas (preanalíticamente), algo que podríamos reconocer sólo si ya comprendiéramos (en parte) el predicado "es verdadero", y (2) que la totalidad de oraciones T fija singularmente la extensión del predicado de verdad. El interés de una teoría de la verdad, vista como una teoría empírica de un lenguaje natural, no consiste en decirnos lo que es la verdad en general, sino en que revela cómo la verdad de cada oración de un *L* particular depende de su estructura y constituyentes.

No tenemos necesidad de preocuparnos, entonces, por el hecho de que una teoría de la verdad no analiza a fondo el concepto preanalítico de verdad. El asunto puede admitirse sin impugnar el interés de la teoría (volveré más adelante a esto). Persiste aún la queja, que también yo admito, de que la teoría no explica o analiza el concepto de referencia. Y esto parece ser una falla atroz, pues mina las pretensiones de la teoría de dar una consideración completa de la verdad de las oraciones.

Esta propiedad de una teoría de la verdad me ha sido señalada con persistencia creciente por cierto número de críticos. Gilbert Harman insiste en el asunto con el objeto de cuestionar la posibilidad de que una teoría de la verdad pueda cumplir alguna tarea en favor de una teoría del significado, tal como yo he venido afirmando.[1] En efecto, dice que sólo puede considerarse que una teoría de la verdad da el significado de las constantes lógicas —da la forma lógica de las oraciones, y hasta ese punto su significado— pero que ella es incapaz de vestir a ese cuerpo desnudo. Hartry Field desarrolla las ideas que he tocado en estas últimas páginas, y concluye que una teoría de la verdad al estilo de Tarski es solamente una parte de una teoría completa.[2] Piensa que además debemos agregar una teoría de la referencia para los predicados y los nombres propios. (He esquematizado su opinión.) Ha habido también críticas conexas provenientes de Kathryn Pyne Parsons, Hilary Putnam y Paul Benacerraf.[3]

Ya he dicho por qué parece que no podemos vivir sin el concepto de referencia; ahora diré por qué pienso que tendríamos que ser reacios a vivir con él. Me interesa lo que considero, históricamente al menos, el problema fundamental de la filosofía del lenguaje, esto es, cómo explicar específicamente conceptos lingüísticos como verdad (de oraciones o emisiones), significado (lingüístico), regla o convención lingüística, nombrar, referir, afirmar, y así por el estilo: cómo analizar algunos o todos

[1] G. Harman, "Meaning and Semantics".
[2] H. Field, "Tarski's Theory of Truth".
[3] K. P. Parsons, "Ambiguity and the Theory of Truth"; H. Putnam, "The Meaning of 'Meaning' "; P. Benacerraf, "Mathematical Truth".

estos conceptos en términos de conceptos de otro orden. Todo acerca del lenguaje puede llegar a parecer enigmático, y lo comprenderíamos mejor si pudiéramos reducir los conceptos semánticos a otros. O si hablar de "reducir" o "analizar" resulta demasiado fuerte (y yo pienso que es así), digamos entonces, tan vagamente como sea posible, comprender los conceptos semánticos a la luz de otros.

"Vivir con" el concepto de referencia significa, en el presente contexto, tomarlo como un concepto a recibir un análisis o interpretación independiente en términos de conceptos no lingüísticos. La cuestión de si la referencia es explícitamente definible en términos de otras nociones semánticas tales como la de satisfacción, o la de caracterizable recursivamente, o no lo es, no es la cuestión esencial: la cuestión esencial es si ella es *el* lugar, o al menos un lugar donde hay contacto directo entre la teoría lingüística y los sucesos, acciones u objetos descritos en términos no lingüísticos.

Si pudiéramos proporcionar el análisis o la reducción que deseamos para el concepto de referencia, a mi parecer se allanaría todo el camino. Habiendo explicado directamente los rasgos semánticos de los nombres propios y de los predicados simples, podríamos pasar a explicar la referencia de términos singulares complejos y de predicados complejos, podríamos caracterizar la satisfacción (como concepto derivativo) y finalmente la verdad. Esta ilustración de cómo hacer semántica es (aparte de los detalles) de vieja data y además muy natural. A menudo se la llama la teoría del Bloque Constructivo. A menudo se la ha ensayado. Y está irremediablemente condenada.

Tenemos que retroceder hasta los primeros empiristas británicos para hallar ejemplos bastante claros de teorías del bloque constructivo. (Berkeley, Hume, Mill.) Los ambiciosos intentos con análisis conductistas del significado realizados por Ogden, por Richards y por Charles Morris no representan casos claros, pues estos autores tendieron a empañar la distinción entre palabras y oraciones ("¡Fuego!" "¡Arbol!" "¡Bloque!") y la mayor pate de lo que dijeron en realidad sólo es aplicable inteligiblemente si se toma a las oraciones como los átomos básicos del análisis. Quine, en el Capítulo II de *Word and Objet*, intenta un análisis conductista, pero aunque su ejemplo más famoso ("Gavagai") es una única palabra, está tratada explícitamente como una oración. Grice, si es que entiendo su proyecto, quiere explicar el significado lingüístico en última instancia apelando a intenciones no lingüísticas, pero a otra vez son los significados de *oraciones*, no de palabras, los que se analizan en términos de alguna otra cosa.

La ilustración histórica, muy simplificada, muestra que a medida que los problemas se fueron haciendo más claros y los métodos más sofisticados, los conductistas y otros que intentaron hacer un análisis radical del lenguaje y la comunicación han ido dejando de lado la aproximación del bloque constructivo en favor de una aproximación que coloca a la oración en el foco de la interpretación empírica.

Y, sin duda, es esto lo que debemos esperar. Las palabras no tienen otra función fuera de desempeñar un papel en las oraciones: sus rasgos

semánticos se abstraen de los rasgos semánticos de las oraciones, tal como los rasgos semánticos de las oraciones se abstraen de *su* participación para ayudar a la gente a alcanzar objetivos o realizar intenciones.

Si el nombre "Kilimanjaro" refiere Kilimanjaro, entonces no hay duda de que existe *alguna* relación entre los hablantes del castellano (o del swahili), la palabra y la montaña. Pero resulta inconcebible que uno tenga que ser capaz de explicar esta relación sin explicar primero el papel de la palabra en las oraciones; y si esto es así, no hay chance de explicar la referencia directamente en términos no lingüísticos.

Es interesante que Quine en el Capítulo II de *Word and Objet* no haga uso del concepto de referencia, ni trate de construirlo. Quine resalta la indeterminación de la traducción y de la referencia. Argumenta que la totalidad de la evidencia disponible para un oyente no determina una forma única de traducir las palabras de un hombre a las de otro; que ni siquiera fija el *aparato* de referencia (términos singulares, cuantificadores e identidad). Pienso que Quine no completa la exposición de su caso. Si es verdad que la evidencia admisible para interpretar un lenguaje ha sido completada cuando sabemos cuáles son las manuales de traducción aceptables de Su lenguaje al Nuestro, luego la evidencia es irrelevante respecto de las cuestiones de referencia y de ontología. Esto responde a que un manual de traducción sólo es un método para pasar de oraciones de un lenguaje a oraciones de otro, y de él nada podemos inferir acerca de las relaciones entre palabras y objetos. Desde luego que sabemos, o creemos saber, qué refieren las palabras en nuestro propio lenguaje, pero esta es información que no se halla en ningún manual de traducción. La traducción es una noción puramente sintáctica. Las cuestiones de referencia no surgen en la sintaxis, y mucho menos se dirimen en ella.

Veamos entonces, en breve, la paradoja de la referencia: Hay dos aproximaciones a la teoría del significado, el método del bloque constructivo, que empieza con lo simple y construye hacia arriba, y el método holístico, que parte de lo complejo (oraciones, en definitiva) y va abstrayendo las partes. El primer método estaría bien si pudiéramos dar una caracterización no lingüística de la referencia, pero no parece haber chances de ello. El segundo comienza del punto (las oraciones) donde podemos tener esperanzas de conectar el lenguaje con la conducta descrita en términos no lingüísticos. Pero parece ser incapaz de dar una consideración completa de los rasgos semánticos de las partes de las oraciones, y sin una consideración así parece que no estamos en condiciones de explicar la verdad.

Volviendo al dilema central: he aquí cómo pienso que se puede resolver. Propongo defender una versión de la aproximación holística, y manifiesto que debemos abandonar el concepto de referencia en tanto concepto básico de una teoría empírica del lenguaje. Reseñaré a grandes rasgos por qué pienso que podemos permitirnos hacer esto.

El argumento contrario al abandono de la referencia expresaba que ella era necesaria para completar una consideración de la verdad. He concedido que una teoría de la verdad al estilo de Tarski no analiza

ni explica el concepto preanalítico de verdad ni el concepto preanalítico de referencia: en el mejor de los casos da la extensión del concepto de verdad para uno u otro lenguaje con un vocabulario primitivo fijo. Pero esto no muestra que una teoría de la verdad absoluta no pueda explicar la verdad de oraciones individuales en base a su estructura semántica; todo lo que ello muestra es que los rasgos semánticos de las palabras no pueden considerarse básicos para la interpretación de la teoría. Lo que se necesita para resolver el dilema de la referencia es la distinción entre explicación *dentro de* la teoría y explicación *de* la teoría. Dentro de la teoría, las condiciones de verdad de una oración se especifican dirigiendo la atención hacia la estructura postulada y conceptos semánticos como los de satisfacción o referencia. Pero cuando se trata de interpretar la teoría como un todo, es la noción de verdad, en tanto aplicada a oraciones cerradas, la que debe ser conectada con los fines y actividades humanos. La analogía con la física es obvia: explicamos los fenómenos macroscópicos postulando una estructura fina no observada. Pero la teoría se pone a prueba a nivel macroscópico. A veces somos lo suficientemente afortunados y encontramos evidencia adicional, o más directa, en favor de la estructura postulada originalmente; pero esto no es esencial para la empresa. Sugiero que las palabras, los significados de las palabras, la referencia y la satisfacción son supuestos que necesitamos para implementar una teoría de la verdad. Ellos sirven a este propósito sin necesidad de una confirmación independiente o de una base empírica.

Ahora tendría que estar aclarado por qué dije, en el párrafo inicial, que una teoría de la verdad del tipo correcto podría ayudar a resolver el aparente dilema de la referencia. La ayuda proviene del hecho de que una teoría de la verdad nos ayuda a responder la pregunta básica de cómo es posible la comunicación por el lenguaje: dicha teoría satisface los dos requerimientos que exigimos de una respuesta adecuada (en el segundo párrafo). Los dos requerimientos se relacionan directamente con la distinción que acabo de hacer entre explicar algo en términos de la teoría, y explicar por qué la teoría es válida (por ejemplo, relacionándola con hechos más básicos).

Tomemos primero la segunda condición. ¿Cómo puede darse una interpretación empírica a una teoría de la verdad absoluta? En el contexto presente resulta esencial que la teoría esté relacionada a conducta y actitudes descritas en términos que no sean específicos del lenguaje u oración involucrado. Una teoría de la verdad al estilo de Tarski proveé el espacio obvio para establecer esta relación: las oraciones T. Si supiéramos que todas ellas son verdaderas, luego una teoría que las implicara satisfaría el requerimiento formal de la Convención T, y daría las condiciones de verdad para cada oración. En la práctica, tendríamos que imaginar al constructor de la teoría suponiendo que algunas oraciones T son verdaderas en base a la evidencia (sea cual fuere), construyendo una teoría idónea, y poniendo a prueba otras oraciones T para confirmar la teoría o proveer una base para su modificación. Una oración T típica, relativizada ahora al tiempo, podría ser:

"Sócrates está volando" es verdadero (en el lenguaje de Smith) en *t* si y sólo si Sócrates está volando en *t*.

Desde el punto de vista empírico, necesitamos una relación entre Smith y la oración "Sócrates está volando" que podamos describir en términos que no se basen en supuestos y que valga cuando y sólo cuando Sócrates esté volando. Damos por supuesto que la teoría contendrá una recursión para un concepto como la satisfacción o la referencia. Mas debemos tratar a estas nociones como construcciones teóricas cuya función se agota en la enunciación de las condiciones de verdad para las oraciones. Ocurre algo similar respecto de la forma lógica atribuida a las oraciones y a toda la maquinaria de términos, predicados, y conectivos y cuantificadore. Nada de esto está abierto a la confrontación directa con la evidencia. Carece de sentido, en esta aproximación, lamentarse de que una teoría da con las condiciones de verdad correctas una y otra vez, pero tiene la forma lógica (o estructura profunda) equivocada. Tendríamos que considerar la referencia de igual modo. Habíamos concordado en que una teoría de este tipo no explica la referencia, al menos en este sentido: no asigna directamente un contenido empírico a las relaciones entre nombres o predicados y objetos. Estas relaciones reciben *indirectamente* un contenido cuando están las oraciones T.

La teoría abandona la referencia, entonces, como parte del costo de volverse empírica. No obstante, no puede decirse que haya abandonado la ontología. Pues la teoría relaciona cada término singular con algún objeto, y dice cuáles entidades satisfacen cada predicado. El arreglárselas sin la referencia nada tiene que ver con abrazar una política que implique arreglárselas sin la semántica o sin la ontología.

No he dicho qué significa contar como evidencia de la verdad de una oración T.[4] La empresa que estamos estudiando se cumple mostrando cómo se puede dar sostén a la teoría por medio de relacionar oraciones T, y nada más, con la evidencia. Está claro que la evidencia, sea cual fuere, no puede describirse en términos que la relacionen de antemano con algún lenguaje particular, y esto sugiere que el concepto de verdad al cual apelamos tiene una generalidad que la teoría no puede pretender explicar.

No se trata de que el concepto de verdad que se usa en las oraciones T pueda ser definido explícitamente en términos no semánticos o reducido a conceptos más conductistas. Reducción y definición son, como dijera al principio, expectativas demasiado grandes. La relación entre teoría y evidencia será seguramente mucho más ligera.

La teoría presupone una noción general y preanalítica de la verdad. Porque tenemos esta noción, podemos determinar qué cuenta como evidencia para la verdad de una oración T. Pero no se requiere lo mismo de los conceptos de satisfacción y referencia. Su papel es teórico, y por lo tanto cuando conocemos cómo operan para caracterizar la verdad cono-

[4] Véase los Ensayos 9-11.

cemos todo lo que se puede conocer acerca de ellos. No necesitamos un *concepto* general de la referencia para la construcción de una teoría adecuada.

No necesitamos el concepto de referencia; tampoco necesitamos la referencia misma, sea ella lo que fuere. Esto es así porque si hay una manera de asignar entidades a las expresiones (una manera de caracterizar la "satisfacción") que produzca resultados aceptables con respecto a las condiciones de verdad de las oraciones, habrá innumerables maneras más de hacerlo igualmente bien. No hay razón, entonces, para llamar "referencia" o "satisfacción" a ninguna de estas relaciones semánticas.[5]

¿Cómo puede una teoría de la verdad absoluta dar una consideración de la comunicación, o ser considerada una teoría del significado? Ella no nos provee de los materiales para definir o analizar frases tales como "significa", "significa lo mismo que", "es una traducción de", etcétera. Es un error pensar que podemos interpretar *automáticamente* que las oraciones T "dan el significado" de las oraciones si como única restricción les aplicamos la exigencia de que resulten verdaderas.

La pregunta que tenemos que formular es si alguien que conoce una teoría de la verdad para un lenguaje L tendría información suficiente como para interpretar lo que dice un hablante de L. Pienso que la manera correcta de investigar esta cuestión es preguntar a su vez si las restricciones empíricas y formales sobre una teoría de la verdad limitan en grado suficiente el rango de teorías aceptables. Supóngase, por ejemplo, que *toda* teoría que satisficiera los requerimientos diera las condiciones de verdad de "Sócrates vuela" tal como se sugiere más atrás. Luego, resulta claro que conocer la teoría (y saber que es una teoría que satisface las restricciones) equivale a conocer que la oración T proporciona *excluyentemente* las condiciones de verdad de "Sócrates vuela". Y esto *es* conocer lo suficiente acerca de su papel en el lenguaje.

Por el momento no puedo imaginar que emerja semejante singularidad. Pero sí pienso que restricciones empíricas razonables sobre la interpretación de oraciones T (las condiciones bajo las cuales las hallamos verdaderas), más las restricciones formales, dejarán las suficientes invariantes entre teorías como para permitirnos decir que una teoría de la verdad captura el rol esencial de cada oración. Una tosca comparación puede ayudarnos a redondear esta idea. Una teoría de la medición de temperatura nos lleva a asignar a los objetos números que miden su temperatura. Tales teorías aplican restricciones formales a esas asignaciones, y además deben sujetarse empíricamente a fenómenos cualitativamente observables. Los números asignados no están determinados únicamente por las restricciones.

Pero el patrón de asignaciones es significativo. (Las temperaturas Fahrenheit y centígrada son transformaciones lineales entre sí; la asig-

[5] Véase J. Wallace, "Only in the Context of a Sentence do Words Have Any Meaning".

nación de números es única en tanto es una transformación lineal.) En forma parecida, sugiero que entre diferentes teorías aceptables de la verdad el significado es invariante. El significado (interpretación) de una oración se da asignando a la oración una posición semántica en el patrón de oraciones que comprende el lenguaje. Diferentes teorías de la verdad pueden asignar diferentes condiciones de verdad a la misma oración (este es el análogo semántico de la indeterminación de traducción de Quine), mientras que las teorías están (casi lo suficientemente) de acuerdo respecto de los roles de las oraciones en el lenguaje.

La idea central es simple. La teoría del bloque constructivo, y las teorías que tratan de dar un contenido rico a cada oración directamente en base a la evidencia no semántica (por ejemplo, las intenciones con que la oración se emite típicamente), tratan de ir muy lejos y muy rápido. La idea presente, en cambio, consiste en aspirar a encontrar un mínimo de información acerca de la corrección de la teoría en cada punto singular; son los infinitos puntos potenciales los que hacen la diferencia. Una teoría fuerte, que se apoya débilmente pero en una cantidad suficiente de puntos, puede producir toda la información que necesitamos acerca de los átomos y las moléculas: en este caso, las palabras y las oraciones.

En pocas palabras: compensamos la escasez de evidencia concerniente a los significados de las oraciones individuales, pero no tratando de producir evidencia para los significados de las oraciones individuales sino teniendo en cuenta la evidencia para una teoría del lenguaje al cual pertenece la oración. Para implementar la teoría necesitamos constructos como las palabras y alguna forma de conectarlas con los objetos. Esta concepción de la forma de elaborar una teoría del significado pertenece esencialmente a Quine. A la idea básica de Quine he sumado la sugerencia de que la teoría tendría que tomar la forma de una teoría de la verdad absoluta. Si efectivamente toma esta forma, podremos recuperar una estructura de las oraciones que esté integrada por términos singulares, predicados, conectivos y cuantificadores, con implicaciones ontológicas del tipo usual. La referencia, sin embargo, quedará en el camino. Ella no desempeña una función esencial en la explicación de la relación entre lenguaje y realidad.

16

La inescrutabilidad de la referencia

La tesis de la inescrutabilidad de la referencia de Quine dice que no hay forma de determinar qué refieren los términos singulares de un lenguaje, o respecto de qué son verdaderos sus predicados, o que al menos no hay forma de determinar esto a partir de la totalidad de la evidencia conductal, real y potencial, y que dicha evidencia es todo lo que interesa a las cuestiones de significado y comunicación. La inescrutabilidad de la referencia surge de consideraciones que se hacen palpables en *Word and Objet*, pero el término y la tesis recién se trataron como temas centrales en "Ontological Relativity". La tesis es importante porque la indeterminación de traducción parte directamente de ella, y resulta más fácil dar fundamentos claros para esta tesis que para algunas otras formas de indeterminación.

Esos fundamentos claros provienen de ejempos: Quine y otros han mostrado cómo construir ejemplos sistemáticos de esquemas alternativos de referencia tales que, si uno de ellos concuerda con toda la evidencia relevante posible, otros lo hacen. Quine argumenta que este hecho tendría que llevarnos a reconocer que la relación de referencia entre objetos y palabras (o sus emisiones) es relativa a una elección arbitraria de un esquema de referencia (o manual de traducción), y que de hecho es relativa a un parámetro básico adicional. El argumento en favor del relativismo me intriga, pues encuentro imposible formular el concepto relativizado de referencia de una manera aceptable. En este trabajo discuto mis dificultades y ofrezco a Quine una salida a una de ellas. Desde luego, espero que dicha salida sea la que Quine ha tenido en mente. En ese caso el ejercicio no será de utilidad para él, pero lo habrá sido para mí.

Para hacer que mi posición general resulte clara desde el principio, acepto la tesis de Quine de la inescrutabilidad de la referencia y en consecuencia también la de la indeterminación de la traducción. Y pienso que acepto ambas principalmente sobre la base de los argumentos que he aprendido de Quine. Pero no veo cómo estos argumentos muestran que la referencia es relativa en la forma que Quine cree que lo es; además, pienso que la visión propia de Quine socava la idea de que la ontología puede ser relativizada.

Donde Quine habla mayormente de relatividad ontológica, yo he venido usando la frase relatividad de referencia. Mis razones para esta modificación se deben a que pretendo concentrarme en un tipo de ejem-

plo de la inescrutabilidad de la referencia particularmente claro y simple. En este tipo de ejemplo se supone que la ontología total es fija, pero la verdad de las oraciones se explica equiparando objetos con palabras de diferentes maneras. Antes de decir más acerca de dichos ejemplos, permítaseme enfatizar el modesto objetivo de este trabajo y clasificar los diversos tipos de indeterminación que discute Quine, pues cualquiera de ellos puede hacer que la referencia sea inescrutable.

Primero, la verdad misma puede ser indeterminada; puede haber un manual de traducción (o mejor, para nuestros propósitos, una teoría de la verdad) para un lenguaje que satisfaga todas las restricciones empíricas relevantes y que determine que cierta oración es verdadera, y otra teoría igualmente aceptable que no determine la verdad de esa oración (por supueto, tendrá que haber también otras diferencias). Luego discutiré brevemente esos casos, pero sólo al pasar.

Segundo, la forma lógica puede ser indeterminada: dos teorías satisfactorias pueden diferir en cuanto a qué consideran términos singulares o cuantificadores o predicados, o inclusive respecto de la misma lógica fundamental.

Tercero, aunque la forma lógica y la verdad sean fijas, teorías aceptables pueden diferir respecto de las referencias que ellas asignan a las mismas palabras y frases. Y aquí podemos subdividir entre casos donde la ontología total difiere y casos donde no o hace. Orientaré mi interés casi exclusivamente hacia la segunda subdivisión del tercer tipo de indeterminación.

Creo que todas estas clases de indeterminación son posibles, pero a excepción de la última, que es la que discutiré, probablemente pienso que el rango de indeterminación es menor que lo que Quine piensa que es. Yo reduciría la primera clase de indeterminación mediante una aplicación del principio de caridad de mucho mayor alcance que aquél que Quine estima esencial. La segunda clase de indeterminación se somete automáticamente a un control mucho mayor si uno insiste, como yo lo hago, en una teoría de la verdad al estilo de Tarski como base de un manual de traducción aceptable. Y, en mi opinión, la variación en la ontología total sólo puede hacerse inteligible en un grado limitado; existe un principio de caridad inversa que juzga que una teoría es mejor cuanto mayor sea la porción de sus recursos propios que lee en el lenguaje para el cual ella es una teoría. No ofreceré aquí las razones de estos puntos de vista, aunque tampoco dependeré de ellos.

Se hace duro, no obstante, hablar de referencia sin colocarla en el contexto de una teoría de la verdad. Ello obedece a que una consideración satisfactoria de la referencia debe asignar una extensión a cada uno de los predicados y términos singulares de un lenguaje, y esto requiere una caracterización recursiva que pueda manejar cuantificadores y tal vez funciones. En cualquier caso, si el lenguaje cuenta con los recursos que nosotros atribuimos con seguridad a un lenguaje natural, querremos introducir un concepto como el de satisfacción que es más general que el de referencia y en base al cual puede definirse la verdad y la referencia (al menos para los predicados). En lo que sigue daré por

supuesto que una relación de referencia vale entre los nombres y lo que se diría que ellos nombran, entre los términos singulares y lo que se diría que ellos denotan, y entre los predicados n-ádicos y los n-tuplos de los que se diría que ellos son verdaderos, estando todos estos tipos de referencia entretejidos entre sí por un conceptom como el de satisfacción que produce una definición de verdad para oraciones cerradas. Todas estas frases "lo que se diría que" son, desde luego, una deferencia para con la inescrutabilidad de la referencia.

La manera más simple y menos cuestionable de mostrar que la referencia es inescrutable se apoya en la idea de una permutación del universo, un mapeo uno a uno de cada objeto con algún otro. Supongamos que $\emptyset$ es tal permutación. Si tenemos un esquema satisfactorio de referencia para un lenguaje que habla de este universo, podemos producir otro esquema de referencia mediante el uso de la permutación: siempre que, en el primer esquema, un nombre refiere un objeto x, en el segundo esquema refiere $\emptyset$ (x); siempre que, en el primer esquema, un predicado refiere (es verdadero para) cada cosa x tal que Fx, en el segundo esquema refiere cada cosa x tal que $F\emptyset$ (x). Entendiendo que la referencia está engranada en cada caso a una caracterización apropiada de una relación como la satisfacción, es fácil ver que las condiciones de verdad que el segundo esquema asigna a una oración serán equivalentes en cada caso a las condiciones de verdad asignadas a esa oración por el primer esquema.[1] Por cierto, como lo señala Wallace, hasta podemos hacer que nuestras dos teorías de la verdad y de la referencia produzcan idénticas condiciones de verdad para todas las oraciones si sumamos a la teoría las suposiciones que nos llevaron a creer que $\emptyset$ era una permutación del universo.

Que yo sepa, los primeros ejemplos del uso de permutaciones del universo para ilustrar la naturaleza escurridiza de la referencia pertenecen a un estudio de Richard Jeffrey.[2] No obstante, su objetivo principal no es la referencia, y sus ejemplos no encajan exactamente dentro de nuestra exposición. Puede encontrarse otros ejemplos en los trabajos de John Wallace y Hartry Field que refiriéramos más atrás. He aquí una simple ilustración: supóngase que cada objeto tiene una y sólo una sombra.[3] Luego podemos entender que $\emptyset$ está expresada por las palabras "la sombra de". En una primera teoría, tomamos el nombre "Wilt" para referir a Wilt y el predicado "es alto" para referir las cosas altas; en la segunda teoría, tomamos "Wilt" para referir la sombra de Wilt y "es alto" para referir las sombras de las cosas altas. La primera teoría nos di-

[1] Para más detalles, y respuestas a diversas dificultades, véase J. Wallace, "Only in the Context of a Sentence do Words Have Any Meaning". Véanse también los dos artículos de Hartry Field, "Quine and the Correspondence Theory" y "Conventionalism and Instrumentalism in Semantics". El presente trabajo ha recibido fuerte influencia de los artículos de Wallace y Field.

[2] R. Jeffrey, reseña de *Logic, Methodology, and the Philosophy of Science.*

[3] Para que el ejemplo funcione correctamente, todo debe ser, al igual que tener, una sombra.

ce que la oración "Wilt es alto" es verdadera si y sólo si Wilt es alto; la segunda teoría dice que "Wilt es alto es verdadera si y sólo si la sombra de Wilt es la sombra de una cosa alta. Las condiciones de verdad son claramente equivalentes. Si no nos molesta hablar de hechos, podríamos decir que el mismo hecho hace que la oración sea verdadera en ambos casos.

Presumo que existen permutaciones del tipo requerido que no necesitan de la ficción. Si queremos concluir que la referencia es inescrutable sin duda también necesitamos suponer que si alguna teoría de la verdad (o de la traducción o de la interpretación) es satisfactoria a la luz de toda la evidencia relevante (real o potencial) luego toda teoría generada a partir de la primera teoría por una permutación será también satisfactoria a la luz de toda la evidencia relevante. Por supuesto que muchos filósofos rechazan este supuesto, pero dado que es una cuestión en la cual concuerdo con Quine, no la defenderé aquí. El punto crucial en el cual coincido con Quine podría expresarse así: toda la evidencia a favor o en contra de una teoría de la verdad (o de la interpretación o de la traducción) se presenta en forma de hechos acerca de qué sucesos o situaciones del mundo causan, o causarían, que los hablantes asintieran o disintieran respecto de cada oración del repertorio del hablante. Probablemente diferimos en algunos detalles. Quine describe los sucesos o situaciones en términos de patrones de estimulación, en tanto que yo prefiero una descripción en términos más parecidos a los de la oración en estudio; Quine daría más peso que yo a una gradación de las oraciones en términos de observacionalidad; y donde él se inclina por el asenso y el disenso a raíz de que ambos sugieren un test conductista, yo reniego del conductismo y acepto las actitudes francamente intensionales hacia las oraciones, tales como considerar verdadero. Hasta donde yo puedo ver, ninguna de estas diferencias interesa al argumento de la inescrutabilidad de la referencia. Lo que interesa es que aquello que causa la respuesta o actitud del hablante es una situación o suceso objetivo, y que la respuesta o actitud está dirigida a una oración o a la emisión de una oración. En tanto nos atengamos a esto, no puede haber una evidencia relevante en base a la cual elegir entre las teorías y sus permutaciones.

La inescrutabilidad de la referencia, entendida y defendida como acabamos de hacerlo, ¿sustenta la idea de que la referencia tendría que relativizarse? Sin duda lo sugiere. No puede satisfacernos una conclusión que nos lleve a aceptar tanto que "Wilt" refiere a Wilt como que "Wilt" refiere$_2$ la sombra de Wilt. Podemos aceptar ambas sin contradicción solamente si ambas pueden ser verdaderas, y está claro que éste no es el caso.

Es posible resolver el problema sin relativizaciones. Todo lo que necesitamos es dejar en claro que "refiere" está siendo usado en dos formas, lo que podríamos indicar con subíndices. "Wilt" refiere a Wilt, y "Wilt" refiere a la sombra de Wilt; esto es una conjunción que puede ser verdadera, y se encuentra bajo las circunstancias que hemos esbozado. Hasta aquí no hemos echado mano a la relativización, aunque el uso de

la misma palabra con diferentes subíndices insinúa una característica común que la relativización podría tornar explícita. Otra razón para inclinarse por la relativización de la referencia es que queremos decir algo como esto: en relación a nuestra primera manera de hacer las cosas, la respuesta correcta a la pregunta acerca de qué refiere "Wilt" es Wilt; en relación a la segunda manera, ella refiere a la sombra de Wilt. Puesto que las maneras de hacer cosas no son entidades atractivas sobre las cuales cuantificar. Quine propone que la referencia sea relativizada a los manuales de traducción. Hartry Field ha señalado que esto no funcionaría.[4] Su argumento destaca que la manera natural de enunciar las condiciones bajo las cuales "*x* refiere *y* en relación a *TM*" es válido es la siguiente: *TM* traduce *x* como "y". Esta sugerencia debe rechazarse pues uno no puede cuantificar entre comillas.

Pienso que hay una razón general por la cual la referencia no puede relativizarse en la forma que lo desea Quine, por lo que resulta inútil tratar de perfeccionar la formulación que acabamos de rechazar. Cuando yo digo "en la forma que lo desea Quine", lo que en realidad estoy objetando no es *cualquier* forma de relativizar la referencia, pues finalmente propondré una forma. Yo objeto la idea de que la referencia pueda relativizarse en una forma tal que fije una ontología. Lo que puedo comprender es la relatividad ontológica. Supongamos que pudiéramos fijar la ontología de "refiere" relativizándola. Entonces habríamos fijado la ontología del lenguaje o hablante al que estaríamos caracterizado mediante el uso de las palabras "refiere". Podría decirse: pero entonces la fijación es sólo relativa a una elección arbitraria. Ninguna evidencia relevante dicta esa elección De aquí la inescrutabilidad. Esta respuesta no da en el centro de la dificultad. Las fijación de referencia y ontología para el lenguaje objeto se ha hecho sobre la base de una elección arbitraria: pero la elección arbitraria tiene éxito en esta tarea sólo si el "refiere" relativizado del metalenguaje ha podido sujetarse de algún modo. Y esto, argumentábamos antes, no puede alcanzarse con ningún lenguaje.

Quizá sea esta línea de razonamiento la que haga decir a Quine que la referencia y la ontología son *doblemente* relativas, una vez a una elección de un manual de traducción y otra vez a alguna teoría o lenguaje de encuadre.[5] Dado que la relativización de la referencia en el metalenguaje no puede restringir la referencia y la ontología para el lenguaje objeto a menos que el predicado de referencia relativizado del metalenguaje tenga una semántica no ambigua. Quine ve una infinita jerarquía de teorías o lenguajes cada uno de los cuales intenta (vanamente) estabilizar el esquema de referencia del lenguaje para el cual proporciona la teoría. Quine compara la relatividad de referencia a una teoría de encuadre con la relatividad de posición en un sistema coordenado. Pero es-

[4] H. Field, "Quine and the Correspondence Theory", 206.

[5] Quine dice a veces "teoría de encuadre", y a veces "lenguaje de encuadre". Véase "Ontological Relativity", 48, 54, 55, 67.

ta comparación fracasa. En el caso de la posición la cuestión es clara. Nunca tiene sentido preguntarse simplemente dónde está un objeto, pero sí tiene sentido preguntar dónde está un objeto en relación a otros objetos: un sistema coordenado (desde luego que la relatividad puede no mencionarse debido a que se presupone algún marco de referencia). La pregunta relativizada ("¿Dónde está la casa de Bronk según el sistema de direcciones del Bronx?) es clara y contestable, y la respuesta es completa. No tiene ningún otro parámetro oculto. Podemos, por supuesto, continuar y formular otra pregunta similar: "¿Dónde está el Bronx?" Y esta pregunta, a su vez, carece de sentido hasta que se la relativiza. Pero una vez relativizada, es clara u contestable. No hay ningún predicado con un espacio extra oculto detrás del predicado de posición relativizado.

Entonces, ¿cómo es posible expresar el relativismo de referencia a un lenguaje de encuadre? Quine admite que "podemos hablar, y efectivamente lo hacemos significativa y distintamente, de conejos y partes, de números y fórmulas", pero sólo en relación a nuestro propio lenguaje.[6] Esto tal vez sugiera que el relativismo puede incluirse en nuestro propio lenguaje. Sin embargo, resulta fácil ver por qué esto es imposible: si la cuestión de la referencia en mi propio lenguaje no está resuelta, no es de ningúna utilidad tratar de decir algo para resolver la cuestión en el mismo lenguaje. Su usted me entiende, interpretará directamente mi emisión de "Wilt es alto". Si tiene dificultades, yo no podré ayudarlo agregando "en castellano", puesto que el mismo problema me forzaría a agregar la misma frase a cada versión sucesivamente aumentada de mi oración. Luego, si la referencia es relativa a mi marco de referencia tal como ya está incorporado en mi propio lenguaje, todo lo que pueda proporcionarse para dar una referencia a mis palabras lo proporciona simplemente la condición de que yo hable mi propio lenguaje. Esto mismo debe ser válido para mi palabra "referencia" cuando se la aplica a otro lenguaje. Pero, otra vez, esto es justamente lo que Quine niega.

Quine compara la relatividad de referencia a un lenguaje de fondo con la situación de la verdad y la satisfacción; también allí, nos recuerda, tenemos la perspectiva de un regreso. Podemos definir la verdad para L en M, pero no en L, podemos definir la verdad para M en M', pero no en M, y así por el estilo. Lo mismo vale para la referencia.

La analogía parece ser defectuosa. Lo que podemos definir en L no puede volverse definible mediante el agregado de un parámetro. Tampoco la verdad en L cuando se la define en M es relativa de algún modo a la verdad M cuando se la define en M'. La verdad es relativa a un lenguaje objeto, pero no a un metalenguaje. El predicado "es verdadero en L" que se presenta en M sí tiene un sentido que podemos, si queremos, especificar todavía en otro lenguaje. Pero, ¿cómo hace esto para que la verdad en L sea relativa a este tercer lenguaje, o aun a M?

Algún concepto del relativismo conceptual parece haber inspirado a Quine a afirmar que la referencia, la verdad y la ontología deben rela-

[6] *Ibídem*, 48.

tivizarse a una teoría o lenguaje de encuadre. Cuando dice que podemos hablar significativa y distintamente de conejos y de partes y que así lo hacemos, pero sólo en relación a nuestro marco de referencia, la explicación de su afirmación, en efecto, no hace otra cosa que decir que cuando hablamos de esta manera debemos hablar un lenguaje que conocemos. Esto, sin embargo, huelga decirlo, y además no promueve ni permite de parte nuestra un enunciado explícito de aquello respecto de lo cual son relativas nuestras afirmaciones. Tal como insiste Quine, podemos enunciar la relatividad si retrocedemos a otro lenguaje, pero si se recurre una vez a esta táctica, tendremos que recurrir a ella cada vez: un regreso infinito. En una situación así, la ontología no es meramente "inescrutable en última instancia"[7], y cualquier pretensión acerca de la referencia, muchas veces relativizada, adolecerá de la misma falta de significado que "Sócrates es más alto que".

Desde luego, Quine está al tanto de la paradoja que se presenta en el relativismo cultural: en todas partes escribe que nadie "puede proclamar el relativismo cultural sin sobrepasarlo, y no puede sobrepasarlo sin dejarlo de lado".[8] Yo diría lo mismo del relativismo ontológico y del relativismo de referencia a una teoría o lenguaje de encuadre. En el ensayo del cual extraje la cita que acabo de transcribir, Quine nos deja nuevamente en el aire en cuanto a la cuestión del relativismo. Argumentar extensamente para convencerse de que puede haber dos teorías tal que ambas impliquen todas las oraciones de observación verdaderas, sean igualmente simples, pero lógicamente incompatibles.[9] De esta manera la verdad es aparentemente relativa a una teoría. Pero Quine se inclina al final por un "franco dualismo" en el cual se usan signos distintivos para anunciar las teorías. Estas resultan ser irreductibles la una a la otra, si bien no contradictorias. Ya no queda relativismo alguno.

En mi opinión la lección consiste en que no podemos encontrar un sentido a una verdad, referencia u ontología relativizadas a una teoría o lenguaje de encuadre. El problema no pasa porque debamos comenzar un regreso que no podemos terminar. El problema es que no comprendemos el primer paso. Permítaseme exponer una vez más mi argumento general al respecto, en una versión ligeramente distinta. Supongamos que *B* enuncia, en sus propias palabras, dos teorías de la verdad para un hablante *A*. Una de sus teorías implica, o especifica, que "Wilt" refiere a Wilt, la otra que "Wilt" refiere la sombra de Wilt. Ahora aparece *C* y trata de producir una teoría para interpretar lo que dice *B*. Naturalmente, descubre muy pronto que la palabra "refiere" debe recibir dos interpretaciones: *C* masculla que *B* tendría que haber usado palabras diferentes, o relativizado de alguna manera el predicado. En cualquier caso, si *C* quiere comprender a *B* debe dar diferentes condiciones de verdad a las dos oraciones de *B* acerca de la palabra "Wilt" de *A*, se preo-

[7] *Ibídem*, 51.
[8] W. V. Quine, "On Empirically Equivalent Systems of the World", 328.
[9] W. V. Quine, "Ontological Relativity", 51.

cupe o no *B* de usar dos palabras o de hacer mención explícita de un parámetro adicional. Pero, ¿puede *C* dar la extensión de los predicados de *B* no arbitrariamente, estén o no éstos subindicados o relativizados? No, porque cualquier teoría satisfactoria que él tenga para comprender a *B* puede transformarse en interminables teorías igualmente satisfactorias. *Nada* de lo que diga *B* puede modificar esto. *A* puede hablar distintamente y significativamente acerca de Wilt y de las sombras. *B* puede hablar distintamente y significativamente Wilt y de las sombras. *B* puede hablar distintamente y significativamente acerca de dos relaciones diferentes entre las palabras de *A* y los objetos. Pero en ningún momento alguien ha sido capaz de especificar de manera singular los objetos para los cuales un predicado es verdadero, no importa cuán arbitrariamente o relativamente lo sea.

Quizás alguien (no Quine) se verá tentado de decir: "Pero al menos el hablante sabe a qué se está refiriendo". Uno debe oponerse firmemente a este pensamiento. Los rasgos semánticos del lenguaje son rasgos públicos. Dada la naturaleza del caso, lo que nadie puede resolver a partir de la totalidad de la evidencia relevante no puede formar parte del significado. Y dado que cada hablante debe saber esto, al menos veladamente, no puede siquiera pretender usar sus palabras con una referencia única, pues sabe que no hay forma de que sus palabras transmitan referencia a otro.

Para recapitular: el argumento de la inescrutabilidad de la referencia tiene dos pasos. En el primer paso reconocemos la equivalencia empírica de esquemas de referencia alternativos. En el segundo paso mostramos que, si bien un intérprete de un esquematizador puede distinguir entre los esquemas de éste, la existencia de esquemas alternativos equivalentes para interpretar al esquematizador impide al intérprete identificar singularmente la referencia de los predicados del esquematizador, en particular su predicado "refiere" (esté o no subindicado o relativizado). Lo que un intérprete no puede decidir respecto de la referencia de las palabras del esquematizador a partir de fundamentaciones empíricas, no puede ser un rasgo empírico de aquellas palabras. Así, dichas palabras no pueden determinar singularmente un esquema de referencia aun cuando se las elija entre alternativas arbitrarias. No se infiere de aquí una relatividad ontológica, pues ella sugiere que cuando se ha tomado una cantidad suficiente de decisiones, arbitrarias o no, la referencia única es posible, y esto es contrario a nuestro argumento de la inescrutabilidad de la referencia.

Si nos fuera posible, mediante alguna elección, arbitraria o no, fijar singularmente la referencia, el terreno se prestaría para que Field (y otros) declararan que la elección no fue arbitraria y que por ello la referencia no es inescrutable. Y esto es, en efecto, lo que ha declarado Field. Según Field, el rol de la referencia no se agota en su contribución a las condiciones de verdad de las oraciones.[10] Como una consecuencia de ello

[10] Véase los dos artículos de H. Field mencionados en la nota 1 al pie de página 231.

encontramos fundamentos para efectuar una elección entre las teorías de la referencia que producen condiciones de verdad equivalentes para las oraciones. Entre esos fundamentos se cuentan ciertas conexiones causales entre los nombres y lo que ellos refieren y (si Putnam está en lo cierto) entre los predicados y aquello para lo cual ellos son verdaderos.

El tópico es amplio e intrincado, y poco es lo que aquí puedo decir de él. Además, no es una cuestión que divida mi punto de vista del de Quine. No obstante, está indudablemente emparentado con la discusión presente. En mi opinión, Field ha confundido el punto al suponer que si hay, digamos, conexiones causales entre palabras y objetos luego no puede darse el caso de que las teorías de la verdad (y del significado y de la referencia) se pongan a prueba tan sólo mediante la evidencia que concierne a las oraciones y sus emisiones. El está fascinado por el hecho de que las teorías estándar de la verdad explican lo que hace verdadera a una oración asignando roles semánticos a las partes. Así, cree él, para que una teoría de la verdad responda a sus afirmaciones y virtudes explicativas debemos ofrecer una consideración independiente de las propiedades semánticas de las partes (de la referencia, del nombre, de la satisfacción). Esta manera de presentar las cosas no consigue vislumbrar la diferencia entre explicar la verdad, dada la teoría, y proporcionar evidencia de que la teoría es verdadera para algún hablante o comunidad. Resulta perfectamente consistente sostener que una teoría sólo puede ponerse a prueba a nivel de las oraciones mientras se explican las características de las oraciones en base a una estructura interna. Pero si esto es así, entonces lo que tendría que verse como más directamente conectado con la evidencia son los rasgos semánticos de las oraciones (por ejemplo la verdad), en tanto que los rasgos semánticos de las palabras, si bien los supuestos cumplirían sus tareas siempre y cuando explicaran los rasgos de las oraciones: abrazar esta postura equivale a suponer que la verdad es más fácil de conectar con la evidencia no lingüística que la referencia. A partir de esto pienso que podemos llegar a algo interesante.[11]

Pero no ahora. El tema que me ocupa ahora es que, aun si las palabras efectivamente tienen, digamos, conexiones causales con lo que ellas refieren, esto no significa que la adecuación de una teoría de la verdad no vaya a ponerse a prueba al nivel oracional. Supongamos que alguna teoría causal de los nombres es verdadera. ¿Cómo vamos a establecer que este hecho es válido para el lenguaje de un hablante o comunidad particular? Sugiero que sólo si encontramos que la teoría causal da cuenta de la conducta lingüística, potencial y real, de los hablantes. Esta conducta está comprometida ante todo con las oraciones y sus emisiones. ¿Por qué esto? Imaginemos que hubiera evidencia de que las pa-

[11] Trato de elaborar esto en el Ensayo 15. Véase también J. Wallace, "Only in the Context of a Sentence do Words Have Any Meaning".

labras, cuando *no* se las emplea en oraciones, tuvieran una conexión causal determinada con objetos pero que este hecho no tuviera relación con la forma en que las palabras se usan en las oraciones. Seguramente concluiríamos que los primeros fenómenos serían irrelevantes en una consideración del lenguaje. Si esto es correcto, la determinación de que una teoría causal de la referencia es verdadera para un hablante debe depender de evidencia extraída a partir de la forma en que se entienden o usan las oraciones, tanto como cualquier otra teoría del lenguaje.

Supongamos que hay conexiones causales entre las palabras, o sus usos, y los objetos. Luego una oración como "Wilt es alto" es verdadera sólo si, entre otras cosas, una emisión de la palabra "Wilt" en este contexto verbal está conectada causalmente con Wilt. ¿Hemos puesto a salvo ahora, al menos parcialmente, al esquematizador de una elección meramente arbitraria de esquemas de referencia para el hablante?

Parece que no, según las razones que explican Wallace y Field. Veamos esto: supóngase, como antes, que ø es una permutación del universo, y que *Cx,y* es una relación causal apropiada entre una palabra y un objeto. Una teoría buena dice que "Wilt" refiere a Wilt sólo si *C* "Wilt", Wilt (estoy haciendo caso omiso a la diferencia entre las palabras y las emisiones de las mismas), mientras que otra teoría empíricamente indistinguible de aquella dice que "Wilt" refiere a ø (Wilt) sólo si *C* "Wilt", ø (Wilt). Desde luego, ningún "refiere" ni "C" puede tener la misma interpretación en ambas teorías, pero sus contrapartes subindicadas son fáciles de especificar. Las dos teorías son claramente distintas, pues "refiere$_2$" no puede tener la misma extensión que "refiere$_2$" y "C_1" no puede tener la misma extensión que "C_2". Pero dado que la primera teoría es satisfactoria, podemos definir "refiere$_2$" y "C_2" en base a "refiere$_1$", a "C_1" y a "ø" para hacer que la segunda teoría sea empíricamente equivalente.

Ninguna teoría causal, ni otro análisis "fisicalista" de la referencia, afectará nuestro argumento en favor de la inescrutabilidad de la referencia, por lo menos mientras admitamos que una teoría satisfactoria es aquella que produce una explicación aceptable de la conducta y disposiciones verbales. Esto obedece a que las restricciones en las relaciones entre referencia y causalidad (o lo que fuera) siempre pueden aprehenderse en forma equivalente mediante vías alternativas de apareamiento de palabras y objetos. El intérprete del esquematizador será capaz, como antes, de determinar que los esquemas del esquematizador son diferentes unos de otros, pero no será capaz de entresacar una única forma correcta de apareamiento de las palabras y objetos del esquematizador. Se sigue que el esquematizador no puede haber usado palabras que determinaran un esquema único. La referencia permanece inescrutable.

Las teorías causales del nombre o de la referencia son por lo general más provocativas, y en alguna versión bien pueden ser verdaderas. De ser así, deben revisarse los conceptos de nombre y de referencia. Pero la cuestión de si una teoría causal de la referencia es verdadera resulta independiente, si yo estoy en lo cierto, de dos cuestiones respecto de las cuales se piensa a menudo que está relacionada: la cuestión de si

una teoría de la verdad debe o no ser puesta a prueba según las reacciones o actitudes de los hablantes hacia las oraciones y la cuestión de la inescrutabilidad de la referencia.

Lo que hemos mostrado, o tratado de mostrar, no es que la referencia no sea relativa sino que hay una forma inteligible de relativizarla que justifique el concepto de relatividad ontológica. La relativización debe aparecer en el lenguaje en el cual aparece el predicado relativizado (de aquí que no pueda aparecer para ese lenguaje o para una teoría de ese lenguaje), y no podemos anunciar que ella resuelva la cuestión de la referencia en ningún lenguaje. Pero *hay* algo a resolver, y la relativización es la única vía atractiva de hacerlo. Por una parte, todos los "esquemas de referencia" que sean aceptables para un hablante o comunidad tienen importantes elementos en común: conducen a condiciones de verdad equivalentes para todas las oraciones, y tal vez conecten la referencia con cuestiones causales de otro tipo según un patrón especificable. Por otra parte, tales esquemas son indistinguibles unos de otros, ciertamente no en base a la evidencia sino en razón de que utilizan predicados que no pueden tener la misma extensión. La suma de estos dos hechos sugiere poderosamente la existencia de un concepto relativizado. Además, está la pregunta a la cual podemos dar una respuesta: "¿Qué esquema está usted usando para dar esa interpretación de las palabras del hablante?" Si yo interpreto (traduzco) que la palabra "Wilt" de un hablante refiere la sombra de Wilt, probablemente necesitaré explicar cómo estoy interpretando su predicado "es alto" de acuerdo a un esquema análogo a mi interpretación de "Wilt". En uno u otro sentido, mi interpretación o traducción es relativa a, o está basada en, un esquema específico. El esquema puede no resolver cuestiones de referencia, pero sí resuelve la forma en que respondo a todas las clases de preguntas acerca de lo que un hablante significa o refiere por medio de una palabra u oración. Frecuentemente Quine presenta el asunto así: "Carece de sentido decir qué son los objetos de una teoría, más allá de decir cómo interpretar o reinterpretar esa teoría en otra,"[12] o también, "Lo que tiene sentido es decir no lo que son los objetos de una teoría, hablando en términos absolutos, sino cómo una teoría de los objetos es interpretable o reinterpretable en otras." La segunda cita me resulta molesta, dado que el "hablando en términos absolutos" sugiere que hay una manera de hablar en términos relativos que decidirá, quizás arbitrariamente, lo que los *objetos* son, y yo he negado esto con vehemencia. Pero no podemos negar que, dado un esquema de interpretación o traducción, hemos decidido qué *palabras* podemos usar en nuestro propio lenguaje para interpretar las palabras de un hablante. ¿Hay una manera inobjetable de señalar la relatividad de nuestra interpretación respecto de nuestro esquema? Pienso que la hay.

La forma en que respondemos preguntas acerca de la referencia es lo único que podemos decir que queda fijado por la relativización, no así la referencia misma. Por eso me parece la forma natural de explicar que

[12] W. V. Quine, "Ontological Relativity", 50.

la relativización explícita que a veces se necesita resulta ser muy familiar: consideramos que el hablante está hablando un lenguaje u otro. Si consideramos que su palabra "conejo" refiere los conejos, consideramos que él está hablando un lenguaje. Si consideramos que su palabra "conejo" refiere cosas que son ø de los conejos, consideramos que está hablando otro lenguaje. Si decidimos cambiar el esquema de referencia, decidimos que él está hablando un lenguaje diferente. En algunos casos la decisión es nuestra; algunos lenguajes son idénticos en cuanto las disposiciones de sus hablantes para emitir oraciones bajo condiciones especificadas son idénticas. No hay forma de determinar cuál de estos lenguajes está hablando una persona.

El punto parecerá menos trivial si reflexionamos que una teoría empírica del lenguaje de una persona no es algo aislado: forma parte de una teoría más general que incluye una teoría de sus creencias, de sus deseos, de sus intenciones, y tal vez más. Si modificamos nuestra interpretación de las palabras de una persona, luego, dada la misma evidencia total, también debemos modificar las creencias y deseos que le atribuimos. No es extraño que podamos considerar que la misma persona está hablando lenguajes diferentes, siempre que podamos efectuar ajustes compensatorios en las otras actitudes que le atribuimos.

El asunto resultará más claro si abandonamos por un momento la restricción que habíamos impuesto al principio para considerar solamente las teorías que mantenían inalterada la verdad de las oraciones. Yo pienso, al igual que Quine, que a menudo se presentan casos en que la totalidad de evidencia relevante en la conducta de una persona es igualmente bien tratada por dos teorías de la verdad, siempre que nuestra teoría efectuemos ajustes compensatorios de sus creencias y otras actitudes, y sin embargo mientras en una teoría una oración particular se interpreta de una forma tal que aparece como verdadera, en la otra teoría ocurre lo contrario. Ian Hacking me presentó una vez este enigma: ¿cómo pueden dos teorías de la verdad ser ambas aceptables si una teoría considera que una determinada emisión es verdadera y la otra no? ¿No es esto una contradicción? No es una contradicción si se relativizan todas las teorías a un lenguaje, tal como sucede con todas las teorías de la verdad. Nuestro error fue suponer que hay un único lenguaje al cual pertenece una emisión dada. Pero podemos considerar que esa emisión pertenece a un lenguaje o a otro, sin caer en paradojas, siempre que dejemos espacio para una modificación en otras partes de nuestra teoría total de una persona.

Hay, entonces, una manera razonable de relativizar la verdad y la referencia: las oraciones son verdaderas, y las palabra refieren, en relación a un lenguaje. Esto puede parecer una cuestión familiar y obvia, y en cierta forma sin duda lo es. Pero hay algunas sutilezas en cuanto a cómo se entiende esto. Por ejemplo, que "Wilt" refiera a Wilt en *L* no es una pretensión empírica, pues si lo fuera tendría que caracterizarse a *L* como *el* lenguaje hablado por alguna persona o personas en un tiempo dado. Tal caracterización no serviría a nuestros propósitos, puesto que admitimos que cuál fuere el lenguaje que habla una persona no es una

cuestión enteramente empírica; la evidencia nos permite una cierta elección de lenguajes, incluso hasta el punto de permitirnos asignar condiciones de verdad contradictorias a una misma oración. Pero aun si consideramos que la verdad es invariante, podemos adecuar la evidencia mediante diversas maneras de aparear palabras y objetos. La mejor manera de anunciar la manera que hemos elegido consiste en nombrar el lenguaje; pero entonces debemos caracterizar el lenguaje como aquél para el cual se ha asignado roles específicos a la referencia, a la satisfacción y a la verdad. Todavía queda una pregunta empírica: ¿es este un lenguaje que la evidencia nos permita atribuir a este hablante?

Lo que nos permite elegir entre varios lenguajes para un hablante es el hecho de que la evidencia —actitudes o acciones dirigidas a oraciones o emisiones— afecta no sólo la interpretación del habla sino también la atribución de creencia, deseos e intenciones (e indudablemente otras actitudes).[13] La evidencia nos consiente una elección entre lenguajes debido a que podemos balancear cualquier elección dada mediante una apropiada elección de creencias y otras actitudes. Esto sugiere una vía más por la que podríamos relativizar una teoría de la verdad o de la referencia: dados determinados supuestos acerca de la naturaleza de la creencia y otras actitudes, podríamos mostrar que, una vez que hemos decidido cuáles son las actitudes de una persona, la elección de un lenguaje ya no está a nuestro alcance. Dada una consideración amplia de la creencia, el deseo, la intención y demás, cuál es el lenguaje que habla una persona pasa a ser una cuestión empírica. Y así tenemos, por fin, una manera bastante sorprendente de comprender en términos significativos la cuestión de lo que refiere una palabra.

Sería un error —espero resulte claro— suponer que de alguna manera podríamos determinar primero lo que una persona cree, desea, pretende y teme y luego pasar a una respuesta definida para la cuestión de lo que sus palabras refieren. Esto es así porque la evidencia de la cual dependen todas estas cuestiones no nos ofrece un camino para separar una por una las contribuciones del pensamiento, la acción, el deseo y la significado. Lo que debemos construir son teorías totales, y muchas teorías funcionarán igualmente bien. Esto equivale a enunciar una vez más la tesis de la inescrutabilidad de la referencia, pero equivale también a insinuar la razón de su existencia.

[13] Se ha puesto énfasis sobre este punto en los Ensayos 10 y 11.

Límites de lo literal

17

Qué significan las metáforas

La metáfora es el sueño del lenguaje y, como todo sueño, su interpretación refleja tanto del intérprete como del originador. La interpretación de los sueños requiere la colaboración entre una persona que sueña y una persona que está despierta, aun cuando sean la misma persona; y el acto de interpretación es en sí mismo una obra de la imaginación. Así también, la comprensión de una metáfora es un empeño tan creativo como la creación de una metáfora, y las reglas que la guían son igualmente escasas.

Excepto en cuestiones de grado, estas observaciones no distinguen la metáfora de las transacciones lingüísticas más rutinarias: toda comunicación por medio del habla supone la interacción de construcción inventiva y comprensión inventiva. La metáfora agrega a lo ordinario un resultado notable que no usa otros recursos semánticos más allá de los recursos de los cuales depende lo ordinario. No hay instrucciones para inventar metáforas; no hay un manual para determinar lo que una metáfora "significa" o "dice"; no hay un test de la metáfora que no requiera el concurso del gusto.[1] Una metáfora implica un tipo y grado de éxito artístico; no hay metáforas que no sean exitosas, tal como no hay bromas que no sean cómicas. Hay metáforas insulsas, pero que igualmente son giros que han sumado algún aporte, aunque no valiera la pena hacerlo o hubiera podido hacerse mejor.

Este trabajo se ocupa de lo que significan las metáforas, y su tesis es que las metáforas significan lo que significan las palabras, en su interpretación más literal, y nada más. Puesto que esta tesis hace frente a los puntos de vista contemporáneos que me son familiares, mucho de lo que tengo que decir será crítico. Pero pienso que la ilustración de la metáfora que emerge una vez que se han disipado el error y la confusión la convierte en un fenómeno más interesante, y no al revés.

El error fundamental que me propongo atacar es la idea de que la metáfora tiene, además de su sentido o significado literal, otro sentido o significado. Esta idea es común a muchos de quienes han escrito acerca de la metáfora: se la encuentra en las obras de críticos literarios como

[1] Pienso que Max Black está equivocado cuando dice: "Las reglas de nuestro lenguaje determinan que algunas expresiones deben contarse como metáforas." ("Metaphor", 29.) No hay tales reglas.

Richards, Empson y Winters; de filósofos desde Aristóteles hasta Max Black; de psicólogos desde Freud y anteriores hasta Skinner y posteriores; y de lingüistas desde Platón hasta Uriel Weinreich y George Lakoff. La idea toma muchas formas, desde la relativamente simple en Aristóteles hasta la relativamente compleja en Black. Aparece en escritos que sostienen que puede obtenerse una paráfrasis literal de una metáfora, pero también la comparten quienes sostienen que típicamente no puede hallarse dicha paráfrasis literal. Muchos ponen el acento en la percepción especial que puede inspirar la metáfora e insisten con vehemencia en afirmar que el lenguaje ordinario, en su funcionamiento usual, no produce tal percepción. Pero también este punto de vista ve a la metáfora como una forma de comunicación paralela a la comunicación ordinaria; la metáfora conduce verdades o falsedades acerca del mundo de manera muy parecida o como lo hace el lenguaje más común, aunque el mensaje pueda considerarse más exótico, más profundo, o más graciosamente ataviado.

El concepto de que la metáfora es fundamentalmente un vehículo para conducir ideas, aunque sean inusuales, me parece tan erróneo como la idea madre de que la metáfora tiene un significado especial. Concuerdo con la opinión de que la metáfora no puede parafrasearse, pero pienso que esto no se debe a que las metáforas dicen algo demasiado nuevo para la expresión literal sino a que no hay nada allí para parafrasear. La paráfrasis, posible o no, es apropiada para lo que se *dice*: tratamos con la paráfrasis de decirlo de otra manera. Pero si yo estoy en lo cierto, una metáfora no dice nada fuera de su significado literal (tampoco quien la construye dice nada fuera de lo literal al usarla). Desde luego, esto no es negar que la metáfora tiene una característica, ni que esa característica puede aprehenderse usando más palabras.

Aquellos que en el pasado han negado que la metáfora tiene un contenido congnitivo además del literal a menudo se han propuesto mostrar que la metáfora es confusa, meramente emotiva, inapropiada para un discurso serio, científico o filosófico. Mis puntos de vista no deben asociarse a esta tradición. La metáfora es un recurso legítimo no sólo en la literatura sino también en la ciencia, la filosofía y el derecho; es efectiva en el elogio y en el abuso, en la plegaria y en la provocación, en la descripción y en la prescripción. Mayormente no estoy en desacuerdo con Max Black, Paul Henle, Nelson Goodman, Monroe Beardsley y otros en cuanto a sus consideraciones de los logros de la metáfora, a excepción de que yo pienso que logra más aun y que sus logros adicionales son de tipo diferente.

Mi desacuerdo gira en torno a la explicación de la forma en que la metáfora genera sus prodigios. Adelantaré algo: mi argumento depende de la distinción entre lo que las palabras significan y el uso que se les da. Pienso que la metáfora pertenece exclusivamente al dominio del uso. Es algo que se obtiene a partir del empleo imaginativo de palabras y oraciones y que depende por completo de los significados ordinarios de esas palabras y por lo tanto de los significados ordinarios de las oraciones que ellas abarcan.

Cuando se explica la forma en que las palabras trabajan en la metáfora no presta ninguna ayuda el supuesto de significados metáforicos o figurativos, o de tipos especiales de verdad poética o metafórica. Estas ideas no explican la metáfora, la metáfora las explica a ellas. Una vez que comprendemos una metáfora podemos llamar "verdad metafórica" a lo que hemos captado y (hasta cierto punto) decir cuál es el "significado metafórico". Pero aflojar simplemente este significado dentro de la metáfora es como explicar por qué un píldora nos hace dormir diciendo que tiene poder dormitivo. Las condiciones de verdad literal y de significado literal pueden asignarse a las palabras y oraciones aparte de los contextos particulares de uso. A esto se debe que el hacer referencia a ellos tenga un poder explicativo genuino.

Trataré de establecer mis ideas negativa acerca de lo que significan las metáforas e introducir mis limitados planteos positivos mediante el examen de algunas teorías falsas de la naturaleza de la metáfora.

Una metáfora nos hace prestar atención a alguna semejanza, a menudo una semejanza nueva o sorprendentemente, entre dos o más cosas. Estas observaciones trilladas pero verdaderas conducen, o parecen conducir, a una conclusión concerniente al significado de las metáforas. Consideremos la semejanza o similitud ordinaria: dos rosas son similares porque comparten la propiedad de ser una rosa; dos niños son similares en virtud de su niñez. O, más simplemente, las rosas son similares porque cada una es una rosa, y los niños, porque cada uno es un niño.

Supóngase que alguien dice "Tolstoy una vez fue un niño". ¿Cómo es que el niño Tolstoy se parece a los otros niños? La respuesta aparece de inmediato: en virtud de exhibir la propiedad de niñez, esto es, reduciendo el palabrerío, en virtud de ser un niño. Si nos cansamos de la frase "en virtud de" podemos, parece, simplificar aun más y decir que el niño Tolstoy comparte con otros niños el hecho de que el predicado "es un niño" se aplica a él; dada la palabra "niño", no tenemos problemas en decir exactamente cómo se parece el niño Tolstoy a otros niños. Podemos hacerlo sin la palabra "niño"; todo lo que necesitamos son otras palabra que signifiquen lo mismo. El resultado final es igual. La semejanza ordinaria depende de agrupamientos establecidos por los significados ordinarios de las palabras. Tal semejanza resulta natural y nada sorprendente, al punto que las maneras familiares de agrupar objetos se vinculan a los significados usuales de las palabras usuales.

Un crítico famoso dijo que Tolstoy fue "un gran niño moralizador". El Tolstoy a que se refiere aquí no es, obviamente, el niño Tolstoy sino el escritor adulto Tolstoy; ésta es la metáfora. Ahora, ¿en qué sentido el escritor Tolstoy es similar a un niño? Quizá lo que vayamos a hacer sea pensar en la clase de objetos que incluye a todos los niños ordinarios y, además, al adulto Tolstoy y luego preguntarnos qué propiedad especial y sorprendente tienen en común los miembros de esta clase. El pensamiento que nos atrae es que con paciencia podríamos llegar a aproximarnos cuanto fuera necesario a la especificación de la propiedad apropiada. En cualquier caso, podríamos alcanzar esa meta perfectamente si encontráramos palabras que significaran exactamente lo que la palabra

metafórica "niño" significa. La cuestión importante, desde mi perspectiva, no es la posibilidad de encontrar o no las demás palabras perfectas sino la presunción que hay algo que debe intentarse, un significado metafórico que hay que igualar. Hasta ahora no he hecho más que esbozar crudamente la forma en que el concepto de significado puede haberse filtrado dentro del análisis de la metáfora, y la respuesta que he sugerido es que, dado que lo que entendemos por semejanza ordinaria acompaña a lo que entendemos por significados ordinarios, resulta natural el supuesto de significados inusuales o metafóricos para que ayuden a explicar las semejanzas que promueve la metáfora.

La idea, entonces, es que en la metáfora determinadas palabras adquieren significados nuevos o, como se los llama a menudo, "extendidos". Cuando leemos, por ejemplo, que "el Espíritu de Dios se dirigió hacia la faz de las aguas", consideraremos que la palabra "faz" tiene un significado extendido (no tomo en cuenta otras metáforas del pasaje). La extensión responde, como puede verse, a lo que los filósofos llaman la extensión de la palabra, esto es, la clase de entidades a las cuales refiere. Aquí la palabra "faz" se aplica a las fases habituales, y además a las aguas.

Sea como fuere, esta consideración no puede ser completa, pues si en estos contextos las palabras "faz" y "niño" son aplicables correctamente a las aguas y al adulto Tolstoy, luego las aguas tienen realmente fases y Tolstoy era literalmente un niño, y todo sentido metafórico se disipa. Si vamos a pensar que las palabras de las metáforas se ocupan directamente de su tarea de aplicación a aquello a lo cual sí son aplicables de manera apropiada, entonces no hay ninguna diferencia entre la metáfora y la introducción de un nuevo término al vocabulario: hacer una metáfora es asesinarla.

Se ha dejado fuera todo recurso al significado original de las palabras. Dependa o no la metáfora de significados nuevos o extendidos, sin duda depende en cierta forma de los significados originales; una consideración adecuada de la metáfora debe admitir que los significados primarios u originales de las palabras permanecen activos en su disposición metafórica.

Quizas, entonces, podamos explicar la metáfora como una especie de ambigüedad: en el contexto de una metáfora, ciertas palabras tienen un sentido sea nuevo, sea ordinario, y la fuerza de ella depende de nuestra incertidumbre mientras oscilamos entre ambos significados. Así, cuando Melville escribe que "Cristo fue un cronómetro", el efecto de la metáfora se produce porque nosotros tomamos primero "cronómetro" en su sentido ordinario y luego en algún sentido extraordinario o metafórico.

Resulta difícil ver cómo esta teoría puede ser correcta, pues la ambigüedad que encierra la palabra, si la hay, se debe al hecho de que en contextos ordinarios ella significa una cosa y en el contexto metafórico significa una cosa distinta; pero en el contexto metafórico no necesariamente dudamos de su significado. Cuando sí dudamos, usualmente se trata de decidir cuál de las intepretaciones metafóricas debemos acep-

tar; rara vez dudamos de que lo que tenemos sea una metáfora. En todo caso, la efectividad de la metáfora sobrevive fácilmente al fin de la incertidumbre sobre la interpretación del pasaje metafórico. Por lo tanto, la metáfora no puede deber su efecto a una ambigüedad de este tipo.[2]

Hay otro tipo de ambigüedad que parece ofrecer una sugerencia mejor. A veces, en un contexto único, una palabra tendrá dos significados y se entenderá que debemos recordar y usar ambos. O, si pensamos que la condición de palabra implica igualdad de significado, podemos considerar que en esta situación lo que parece ser una palabra única son en efecto dos. Cuando Shakespeare describe la obscena bienvenida que recibe Cressida en el campamento griego, Néstor dice, "Nuestro general os saluda con un beso". Aquí tenemos que considerar "general" de dos maneras: una vez aplicada a Agamenón, quien es el general; y otra vez, puesto que ella está besando a todos, no la aplicamos a nadie en particular, sino a todos en general. En realidad tenemos la conjunción de dos oraciones: nuestro general, Agamenón, os saluda con un beso; y todos en general os saludamos con un beso.

Esto es un recurso legítimo, un juego de palabras, pero no es el mismo recurso de la metáfora, pues en ella no hay una necesidad esencial de reiteración; cualesquiera fueren los significados que asignemos a las palabras, ellos se mantienen a través de cada lectura correcta del pasaje.

Una modificación plausible de esta última sugerencia sería considerar que la palabra (o las palabras) clave de una metáfora tiene dos tipos diferentes de significado inmediato, uno literal y uno figurativo. Imaginemos que el significado literal está latente, es algo que percibimos, que puede afectarnos sin afectar el contexto, mientras que significado figurativo carga con el peso efectivo. Y por último, debe haber una regla que conecte los dos significados, pues de lo contrario la explicación degenera en una forma de la teoría de la ambigüedad. La regla, al menos para muchos casos típicos de metáfora, dice que en su rol metafórico la palabra se aplica a todo aquello a lo que se aplica en su rol literal, y algo más.[3]

Esta teoría puede parecer compleja, pero es sorprendentemente similar a la que propuso Frege para considerar la conducta de los términos referentes en las oraciones modales y en las oraciones acerca de

[2] Nelson Goodman dice que la metáfora y la ambigüedad difieren principalmente "en que los varios usos de un término meramente ambiguo son contemporáneos e independientes", mientras que en la metáfora "un término con una extensión establecida por hábito se aplica en todas partes bajo la influencia de ese hábito"; este autor sugiere que como nuestro sentido de la historia de los "dos usos" en la metáfora se desvanece, la palabra metafórica se torna meramente ambigua (*Languages of Art*, 71). En efecto, en muchos casos de ambigüedad un uso surge del otro (como dice Goodman) y por lo tanto no pueden ser contemporáneos. Pero el error básico, que Goodman comparte con otros, está en la idea de que la metáfora involucra dos "usos" en una forma que nada tiene que ver con la que tienen en la ambigüedad.

[3] La teoría que se describe es esencialmente la de Paul Henle, "Metaphor".

actitudes proposicionales como creencia y deseo. Según Frege, cada término referente tiene dos (o más) significados, uno que fija su referencia en contextos ordinarios y otro que fija su referencia en los contextos especiales creados por los operadores modales o los verbos psicológicos. La regla que conecta ambos significados puede expresarse así: el significado de la palabra en los contextos especiales hace que la referencia en dichos conextos sea idéntica al significado en los contextos ordinarios.

Ha aquí la ilustración completa, poniendo a Frege junto con un punto de vista fregeano de la metáfora: pensaremos que una palabra tiene, además de su campo de aplicación o referencia mundano, dos campos de aplicación especiales o supermundanos, uno para la metáfora y otros para los contextos modales y demás. En ambos casos el significado original se mantiene para desempeñar su tarea en virtud de una regla que relaciona los diversos significados.

Habiendo acentuado la posible analogía entre el significado metafórico y los significados fregeanos para contextos oblicuos, me ocuparé de una dificultad tremenda para el mantenimiento de la analogía. Imaginemos que estoy ocupado tratando de enseñar a un visitante de Saturno el uso de la palabra "suelo". Echo a mano a recursos familiares, lo llevo de un suelo a otro, señalando y golpeando con el pie y repitiendo la palabra. Lo induzco a hacer experimentos, tocando los objetos tentativamente con su tentáculo mientras le indico sus aciertos y errores. Quiero que él no sólo logre aprender que estos objetos o superficies particulares son suelos, sino también pueda distinguir un suelo cuando lo tenga a la vista o al alcance del tacto. Los datos que estoy aportando no le *dicen* a él lo que necesita saber, sino que con suerte lo ayudan a aprenderlo.

¿Cómo debemos llamar a este proceso, aprender algo acerca del mundo o aprender algo acerca del lenguaje? Es una pregunta extraña, pues lo que se aprende es que una porción del lenguaje refiere una porción del mundo. Pero todavía resulta fácil distinguir entre la tarea de aprender el significado de una palabra y usar la palabra una vez que se ha aprendido su significado. Si comparamos estas dos actividades, es natural decir que la primera supone aprender algo acerca del lenguaje, en tanto que la segunda es típicamente aprender algo acerca del mundo. Si mi saturniano ha aprendido a usar la palabra "suelo", puedo tratar de decirle algo nuevo, que *aquí* hay un suelo. Si ya domina el arte de la palabra, le habré dicho algo acerca del mundo.

Mi amigo de Saturno me transporta ahora través del espacio, llegamos a su esfera doméstica, y mirando hacia atrás hacia la remota Tierra le digo, señalándola, "suelo". Quizá piense que esto es aún parte de la lección y suponga que la palabra "suelo" se aplica apropiadamente a la Tierra, al menos cuando se la ve desde Saturno. Pero, ¿y si yo pensara que él ya conoce el significado de "suelo", y estuviera recordando cómo Dante, desde un lugar similar en los cielos, veía la Tierra habitada como "el pequeño suelo redondo que nos hace apasionados"? Mi propósito fue la metáfora, y no un ejercicio de uso del lenguaje. ¿Qué diferencia tendrá para mi amigo entenderlo de una u otra forma? Según la teoría de la metáfora que estamos considerando, muy poca diferencia, pues de

acuerdo con esa teoría una palabra tiene un significado nuevo en el contexto metafórico; la ocasión de la metáfora sería entonces la ocasión para aprender el nuevo significado. Tendríamos que ponernos de acuerdo en que en ciertos aspectos hay una diferencia relativamente pequeña si, en un conexto dado, pensamos que una palabra está siendo usada metafóricamente o en una forma previamente desconocida pero literal. Empson, en *Some Versions of Pastoral*, cita estas líneas de Donne: *"As our blood labours to beget / Spirits, as like souls as it can,... / So must pure lover's soules descend. ..."* El lector moderno está casi seguro, señala Empson, de tomar la palabra "spirits" metafóricamente en este pasaje, como si fuera aplicable sólo por extensión a algo espiritual. Pero para Donne no había metáfora. Escribe este autor en sus *Sermons*, *"The Spirits... are the thin and active parts of the blood, and are a kind of middle nature, between soul and body."* Saber esto no interesa mucho; Empson tiene razón cuando dice, "Es curioso cómo los cambios en el mundo [esto es, en lo que nosotros pensamos que significa] dejan indemne a la poesía."[4]

Este cambio puede ser, en algunos casos al menos, difícil de apreciar, pero a menos que haya un cambio, la mayor parte de lo que se considera interesante acerca de la metáfora se pierde. He estado destacando esto por medio del contraste entre el aprendizaje de un nuevo uso para una palabra vieja y el uso de una palabra que ya se comprende; dije que en un caso nuestra atención se dirige al lenguaje, y en el otro a aquello de que trata el lenguaje. Sugerí que la metáfora pertenece a la segunda categoría. Esto también puede advertirse si se considera metáforas muertas. Alguna vez, supongo, los ríos y las botellas no tenían literalmente bocas, como las tienen hoy. Si pensamos en el uso presente, no interesa si consideramos que la palabra "boca" es ambigua pues se aplica a las entradas de los ríos y las aberturas de las botellas al igual que a las aberturas de los animales, o si pensamos que hay un único campo amplio de aplicación que abarca a ambas. Lo que sí interesa es que cuando "boca" sólo se aplicaba metafóricamente a la botellas, la aplicación hizo que el oyente *notara* una similitud entre las aberturas de los animales y las de las botellas. (Considérese la referencia que Homero hace a las heridas como bocas.) Una vez que uno posee el uso presente de la palabra, con su aplicación literal a las botellas, ya no queda nada para notar. No hay que buscar una semejanza pues ella consiste simplemente en ser referida por la misma palabra.

La novedad no es la clave. En su contexto una palabra que una vez se consideró metáfora continúa siendo metáfora después de escuchársela cien veces, en tanto que otra palabra puede apreciarse con facilidad en un nuevo rol literal la primera vez que se la encuentra. Lo que entendemos por elemento de novedad o de sorpresa en una metáfora es una característica estética incorporada que podemos experimentar una

[4] W. Empson, *Some Versions of Pastoral*, 133.

y otra vez, como la sorpresa al escuchar la Sinfonía Nº 94 de Haydn o una engañosa cadencia familiar.

Si la metáfora involucrara un segundo significado, tal como lo hace la ambigüedad, podríamos pretender ser capaces de especificar el significado especial de una palabra en una disposición metáforica con sólo esperar hasta que la metáfora muriera. El significado figurativo de la metáfora viva quedaría inmortalizado en el significado literal de la muerta. Pero aunque algunos filósofos han sugerido esto, la idea parece completamente equivocada. "El está ardiendo" es genuinamente ambigua (dado que puede ser verdadera en un sentido y falsa en otro), pero si bien esta frase familiar es indudablemente el cuerpo de una metáfora, "El está ardiendo" ya no sugiere hoy que él está muy enojado. Cuando la metáfora estaba en actividad, nos lo habríamos representado con fuego en los ojos y con humo saliéndole de las orejas.

Podemos aprender mucho acerca de lo que significan las metáforas si las comparamos con los símiles, pues un símil nos dice, en parte, qué nos induce ligeramente la metáfora a hacernos notar. Supongamos que Goneril hubiera dicho, pensando en Lear: "Los necios son como bebés de nuevo"; luego habría usado esas palabras para afirmar una similitud entre los necios y los bebés. Lo que ella efectivamente dijo, desde luego, fue "Los necios son bebés de nuevo", usando las palabras de esta manera para indicar lo que declaraba el símil. Esta manera de pensar puede inspirar otra teoría del significado figurativo o especial de las metáforas: el significado figurativo de una metáfora es el significado literal del símil correpondiente. Así "Cristo fue un cronómetro" en su sentido figurativo es sinónimo de "Cristo fue como un cronómetro", y el significado metafórico que una vez encerraba "El está ardiendo" se ve liberado en "El está como alguien que está ardiendo" (o tal vez "El está como ardiendo").

Existe también, con seguridad, la dificultad de identificar el símil que corresponde a una metáfora dada. Virginia Woolf dijo que un erudito es "un hombre o mujer con una inteligencia de pura sangre que conduce su mente al galope a campo traviesa en persecución de una idea". ¿Qué símil le corresponde? Quizás algo como esto: "Un erudito es un hombre o mujer cuya inteligencia es como un caballo de pura sangre y que insiste en pensar acerca de una idea como un jinete que galopa a campo traviesa en persecución de... bueno, algo."

La idea de que significado especial de una metáfora es idéntico al significado literal de un símil correspondiente (como quiera que "correspondiente" se interprete) no tendría que confundirse con la teoría común de que una metáfora es un símil elíptico.[5] Esta teoría no hace una distinción de significado entre una metáfora y algún símil conexo y no prepara un terreno para hablar de significados figurativos, metafóricos

[5] J. Middleton Murray dice que una metáfora es un "símil condensado" (*Countries of the Mind*, 3.). Max Black atribuye una idea similar a Alexander Bain, *English Composition and Rhetoric.*

o especiales. Es una teoría que gana sin discusión en lo que a simplicidad se refiere, pero que al mismo tiempo parece demasiado simple para funcionar. Puesto que si hacemos que el significado literal de la metáfora sea el significado literal de un símil corespondiente, no hacemos un lugar para lo que habíamos considerado originalmente como el significado literal de la metáfora, cuando habíamos acordado casi desde el principio que *este* significado era esencial para el funcionamiento de la metáfora, al margen de cualquier otra cosa que tuviera que incorporarse en la forma de un significado no literal.

Tanto la teoría del símil elíptico de la metáfora como su variante más sofisticada, que iguala el significado figurativo de la metáfora con el significado literal de un símil, comparten un defecto fatal. Ambas hacen que el significado oculto de la metáfora resulte demasiado obvio y accesible. En cada caso el significado oculto se halla tras una simple mirada al significado literal de lo que usualmente es un símil dolorosamente trivial. Esto es como aquello: Tostoy como un niño, la Tierra como un suelo. Resulta trivial porque todo es como todo, interminablemente. Las metáforas son a menudo muy difíciles de interpretar y, se dice, imposibles de parafrasear. Pero con esta teoría, la interpretación y la paráfrasis están siempre al alcance del más inexperto.

Pienso que estas teorías del símil se han hallado aceptables sólo porque se las ha confundido con una teoría bastante diferente. Consideremos esta observación de Max Black:

> Cuando Schopenhauer dijo que una prueba geométrica era una trampa para ratones, estaba *diciendo* (si bien no explícitamente), de acuerdo con tal idea: "Una prueba geométrica es *como* una trampa para ratones, puesto que ambas ofrecen una recompensa engañosa, tientan a sus víctimas gradualmente, llevan a una sorpresa desagradable, etcétera." Esta es una idea de la metáfora como *símil* condensado o elíptico.[6]

Aquí pudo discernir dos confusiones. Primero, si las metáforas son símiles elípticos, dicen *explícitamente* lo que son los símiles, pues la elipsis es una forma de abreviación y no una paráfrasis o indirecta. Pero, y este es el punto más importante, el enunciado de lo que dice la metáfora según Black va mucho más allá de cualquier cosa que pueda dar el símil correspondiente. El símil simplemente dice que una prueba geométrica es como una trampa para ratones. No nos *cuenta* ya qué similitudes podemos notar que hace la metáfora. Black menciona tres similitudes, y naturalmente nosotros podemos continuar siempre aumentando la lista. Pero, ¿se supone que esta lista, una vez revisada y aumentada en la forma correcta, dará el significado *literal* del símil? Es seguro que no, puesto que el símil no declara nada más que la similitud. Si se supone que la lista proporciona el significado figurativo del símil, entonces no aprendemos nada acerca de la metáfora a partir de la comparación con el símil: sólo que ambos tienen el mismo significado figura-

[6] M. Black, "Metaphor", 35.

tivo. Nelson Goodman afirma en efecto que "la diferencia entre el símil y la metáfora es despreciable", y prosigue, "Sea la locución 'es como' o 'es', la figura *asemeja* la ilustración con la persona mediante la selección de un determinado rasgo común..."[7] Goodman está considerando la diferencia entre decir que una ilustración es triste y decir que ella es como una persona triste. Es claramente verdadero que ambos dichos asemejan la ilustración con una persona, pero a mí me parece un error sostener que cualquiera de las dos formas de hablar "selecciona" un determinado rasgo común. El símil nos dice que hay una semejanza y nos deja la tarea de seleccionar algún rasgo común (o más de uno); la metáfora no afirma explícitamente que hay una semejanza, pero si la aceptamos como metáfora, nos vemos otra vez llevados a buscar rasgos comunes (no necesariamente los mismos rasgos que sugiere el símil conexo; pero este es otro tema).

Precisamente porque un símil guarda una declaración de similitud en la manga, pienso que es mucho menos plausible que en el caso de la metáfora sostener que hay un segundo significado oculto. En el caso del símil, notamos lo que él dice literalmente, esto es, que dos cosas se parecen entre sí; luego observamos los objetos y consideramos, dado el contexto, qué similitud correspondería al caso. Una vez decidido esto, podríamos decir luego que el autor del símil pretendía —es decir, se propondría— que nosotros advirtiéramos la similitud. Pero habiendo apreciado la diferencia entre lo que las palabras significaban y lo que el autor logró gracias al uso de dichas palabras, no tendríamos que vernos tentados de explicar lo que ha sucedido dotando a las palabras mismas de un segundo sentido o sentido figurativo. La clave del concepto de significado lingüístico pasa por explicar lo que puede hacerse con palabras. Pero el supuesto significado figurativo de un símil no explica nada; no es un rasgo de la palabra que la palabra tiene anteriormente e independientemente del contexto de uso, ni descansa sobre costumbre lingüística alguna, excepto aquellas que gobiernan el significado ordinario.

Lo que las palabras sí hacen con su significado literal en el símil también debe ser factible para ellas en la metáfora. Una metáfora dirige la atención a los mismos tipos de similitud, si no a las mismas similitudes, hacia los cuales la dirige el símil correspondiente. Pero entonces los paralelos y analogías inesperados o sutiles que la metáfora debe organizar no necesitan depender, para su organización, de otra cosa fuera de los significados literales de las palabras.

La metáfora y el símil no son más que dos recursos entre los innumerables recursos que sirven para ponernos al tanto de distintos aspectos del mundo invitándonos a hacer comparaciones. Citaré unas pocas estrofas de "The Hippopotamus" de T. S. Eliot:

The broad-backed hippopotamus
Rests on his belly in the mud;

[7] N. Goodman, *Languages of Art*, 77-8.

Althouth he seems so firm to us
He is merely flesh and blood.

Flesh and blood is weak and frail,
Susceptible to nervous shock;
While the True Church can never fail
For it is based upon a rock.

The hippo's feeble steps may err
In compassing material ends,
While the True Church need never stir
To gather in its dividens.

The 'potamus can never reach
The mango on the mango-tree;
But fruits of pomegranate and peach
Refresh the Church from over sea.[8]

Aquí no se nos dice que la Iglesia parezca un hipopótamo (como en el símil) ni se nos incita a hacer esta comparación (como en la metáfora), pero no cabe duda de que se usan las palabras para dirigir nuestra atención hacia las similitudes entre ambas cosas. Tampoco tenemos que sentirnos muy inclinados, en este caso, por el supuesto de significados figurativos, pues ¿sobre qué palabras u oraciones los haríamos descansar? El hipopótamo realmente descansa en el barro sobre su panza; la Iglesia, dice literalmente el poema, nunca puede fallar. Por supuesto, el poema insinúa muchas cosas que van más allá del significado literal de las palabras, pero insinuación no equivale a significado.

Hasta ahora el argumento nos ha orientado hacia la conclusión de que tanto como pueda explicarse de la metáfora en términos de significado puede, y por cierto debe, ser explicado recurriendo a los significados literales de las palabras. Una consecuencia de esto es que las oraciones en las cuales se presentan las metáforas son verdaderas o falsas en una forma normal y literal, puesto que si las palabras que hay en ellas no tienen significados especiales, las oraciones no tiene una verdad especial. Esto no es negar que exista una cosa como la verdad metafórica, sino que es sólo negarlo respecto de las oraciones. La metáfora nos lleva efectivamente a advertir lo que de otro modo podría pasar inadvertido, y no hay razón, supongo yo, para no decir que estas visiones, pensamientos, y sentimientos inspirados por la metáfora son verdaderos o falsos.

Si una oración usada metafóricamente es verdadera o falsa en el sentido habitual, luego está claro que ella es usualmente falsa. La diferencia semántica más obvia entre el símil y la metáfora consiste en que todos los símiles son verdaderos y la mayor parte de las metáforas es falsa. La Tierra es como un suelo, el asirio se precipitó como un lobo so-

[8] T. S. Eliot, *Selected Poems.*

bre el rebaño, porque todo es como todo. Pero convierta estas oraciones en metáforas, y las volverá falsas; la Tierra es como un suelo, pero no es un suelo; Tolstoy, ya adulto, era como un niño, pero no era un niño. De ordinario usamos el símil sólo cuando sabemos que la metáfora correspondiente es falsa. Decimos que el señor S. es como un cerdo pues sabemos que no es un cerdo. Si hubiéramos usado la metáfora diciendo que es un cerdo, esto no habría sido porque cambiamos nuestra idea acerca de los hechos sino porque elegimos llegar a la idea por otra vía.

No interesa la falsedad real, sino que la oración pueda ser considerada falsa. Nótese qué ocurre cuando una oración que usamos como metáfora, creyéndola falsa, pasa a entenderse como verdadera debido a un cambio en lo que se cree acerca del mundo. Cuando se informó que el avión de Hemingway había sido avistado, destruido, en Africa, el periódico *Mirror* de Nueva York publicó un titular en primera plana que decía "Hemingway perdido en Africa", donde la palabra "perdido" se usaba para sugerir que estaba muerto. Cuando resultó ser que estaba vivo, el *Mirror* hizo que el titular fuera interpretado literalmente. Considérese también este caso: una mujer se ve con un hermoso vestido y exclama "¡Este vestido es un sueño!" y en ese momento se despierta. El punto de la metáfora es que el vestido es como un vestido que uno soñaría y por lo tanto no es un vestido de ensueño. Henle ofrece un buen ejemplo de *Antonio y Cleopatra* (2.2):

> *The barge she sat in, like a burnish'd throne,*
> *Burn'd on the water*

Aquí el símil y la metáfora interactúan raramente, pero la metáfora se disiparía si se imaginara una conflagración literal. De una manera muy similar el efecto usual de un símil puede sabotearse si se toma la comparación muy seriamente. Woody Allen escribe: "El juicio, que tuvo lugar durante las semanas siguientes, fue como un circo, aunque hubo dificultades para meter a los elefantes en la sala del tribunal".[9]

Por lo general, sólo cuando se considera que una oración es falsa la aceptamos como una metáfora y comenzamos a perseguir la implicación oculta. Probablemente sea por esta razón que las oraciones más metafóricas son *patentemente* falsas, tal como todos los símiles son trivialmente verdaderos. En una oración metafórica, el absurdo o la contradicción garantizan que no creeremos en ella y nos invitan, bajo las circunstancias apropiadas, a tomar la oración metafóricamente.

El caso usual de la metáfora es la falsedad patente, pero en ocasiones también servirá la verdad patente. "Negocios son negocios" es demasiado obvia en su sentido literal como para que se considere que se la ha emitido para conducir información, por lo que buscamos otro uso en ella; Ted Cohen nos recuerda, en la misma línea, que ningún hombre es una

[9] Woody Allen, "Condemned".

isla.[10] La clave es la misma. El significado ordinario en el contexto de uso es lo suficientemente extraño como para urgirnos a dejar de lado la cuestión de la verdad literal.

Ahora me permitiré hablar de un tema en cierto modo platónico, comparando el hacer una metáfora con el decir una mentira. La comparación es pertinente porque mentir, como hacer una metáfora, no concierne al significado de las palabras sino a su modo. Se dice a veces que el decir una mentira implica lo que es falso; pero esto es un error. Para decir una mentira se requiere, no que lo que uno dice sea falso, sino que uno piense que es falso. Dado que usualmente creemos en las oraciones verdaderas y no creemos en las falsas, casi todas las mentiras son falseadas; pero en todo caso particular esto es un accidente. El paralelo entre hacer una metáfora y decir una mentira se ve enfatizado por el hecho de que una misma oración puede usarse para ambos propósitos sin que cambie su significado. Así, una mujer que creyera en las brujas pero que no pensara que su vecina es una bruja podría decir, "Ella es una bruja" con una intención metafórica; la misma mujer, aun creyendo lo mismo respecto de las brujas y de su vecina pero con la intención de engañar, podría usar las mismas palabras en busca de un efecto muy diferente. Puesto que oración y significado son iguales en ambos casos, a veces es difícil probar cuál es la intención detrás de su emisión; de esta manera, un hombre que dice "Lattimore es comunista" y pretende mentir, siempre puede evadirse argumentando una metáfora.

Lo que hace la diferencia entre una mentira y una metáfora no es la diferencia entre las palabras usadas o lo que ellas significan (en un sentido estricto de significado), sino cómo se usan las palabras. Usar una oración para decir una mentira y usarla para hacer una metáfora son, desde luego, usos completamente diferentes, tan diferentes que no interfieren uno con el otro, como lo hacen por ejemplo actuar y mentir. Al mentir, uno tiene que hacer una aserción de manera de representarse creyendo lo que uno cree; al actuar, la aserción queda excluida. A la metáfora no le interesa esta diferencia. Puede ser un insulto, y por lo tanto ser una aserción, decirle a un hombre "Usted es un cerdo". Pero no había una metáfora de por medio cuando (supongamos) Odiseo empleó las mismas palabras respecto de sus compañeros en el palacio de Circe; una historia, sin duda, y por ello no una aserción; pero la palabra, siquiera una vez, fue usada literalmente para los hombres.

Ninguna teoría del significado metafórico o de la verdad metafórica puede ayudar a explicar cómo funciona la metáfora. La metáfora corre por los mismos carriles lingüísticos familiares que siguen las más simples oraciones; esto lo vimos al considerar el símil. Lo que distingue a la metáfora no es su significado sino su uso: en esto es como la aserción, la insinuación, la mentira, la promesa o la crítica. Y el uso especial

[10] T. Cohen, "Figurative Speech and Figurative Acts", 671. Dado que la negación de una metáfora siempre parece ser una metáfora potencial, puede haber tantas trivialidades entre las metáforas potenciales como absurdos hay entre las reales.

que asignamos al lenguaje en la metáfora no es —no puede ser— "decir algo" especial, por indirecto que fuera. Pues la metáfora *dice* sólo lo que tiene a la vista: usualmente una falsedad patente o una verdad absurda. Y esta verdad o falsedad manifiestas no necesitan paráfrasis: su significado está dado en el significado literal de las palabras.

¿Qué vamos a hacer, entonces, con la incesante energía que se ha gastado, y se sigue gastando, en métodos y recursos para extraer el contenido de la metáfora? Los psicólogos Robert Verbrugge y Nancy McCarrell nos dicen que:

> Muchas metáforas atraen la atención hacia sistemas comunes de relaciones o transformaciones comunes, en los cuales la identidad de los participantes es secundaria. Por ejemplo, consideremos las oraciones: *Un auto es como un animal. Los troncos de los árboles son pajillas para las hojas y ramas sedientas.* La primera oración dirige la atención hacia sistemas de relaciones entre el consumo de energía, la respiración, el movimiento autoinducido, los sistemas sensoriales y, posiblemente, un homúnculo. En la segunda oración, la semejanza es de un tipo más restringido de transformación: la succión de fluido a través de un espacio cilíndrico orientado verticalmente de una fuente de fluido a un destino.[11]

Verbrugge y McCarrell no creen que haya una línea muy tajante entre los usos literales y metafóricos de las palabras; ellos piensan que muchas palabras tienen un significado "difuso" que queda fijado, si ello ocurre, por el contexto. Pero seguramente esta difusión, aunque se la ilustre y se la explique, no puede borrar la línea entre lo que una oración significa: literalmente (dado su contexto) y aquello hacia lo que ella "atrae nuestra atención" (dado su significado literal tal como lo fija el contexto). El pasaje que he citado no emplea una tal distinción: aquello hacia lo que —según éste— las oraciones ejemplificadoras dirigen nuestra atención son hechos expresados por paráfrasis de las oraciones. Verbrugge y McCarrell simplemente quieren insistir en que una paráfrasis correcta puede enfatizar "sistemas de relaciones" más que semejanzas entre objetos.

Según la teoría de la interacción de Black, una metáfora nos hace aplicar un "sistema de lugares comunes" asociado con la palabra metafórica al sujeto de la metáfora: en "El hombres es un lobo" aplicamos atributos que son lugares comunes (estereotipos) del lobo al hombre. De esta manera, dice Black, la metáfora "selecciona, enfatiza, suprime, y organiza características del sujeto principal mediante la implicación de enunciados acerca del mismo que normalmente se aplican al sujeto subsidiario".[12] Según Black, si la paráfrasis fracasa ello no se debe a que la metáfora no tiene un contenido cognitivo especial, sino a que la paráfrasis "no tendrá el mismo poder para informar e iluminar que el original...

[11] R. R. Verbrugge y N. S. McCarrell, "Metaphoric Comprehensión: Studies in Reminding and Resembling", 499.
[12] M. Black, "Metaphor", 44-5.

Uno de los puntos que más deseo acentuar es que la pérdida en esos casos es una pérdida de contenido cognitivo; la debilidad relevante de la paráfrasis literal no proviene de que ella pueda ser extensa hasta el tedio o explícita hasta aburrir; no logra ser una traducción pues no logra ofrecer la percepción que ofrecía la metáfora."[13]

¿Cómo puede esto ser correcto? Si una metáfora tiene un contenido cognitivo especial, ¿por qué tendría que ser tan difícil o imposible ponerlo al descubierto? Si, como afirma Owen Barfield, una metáfora "dice una cosa y significa otra", ¿por qué ocurriría que cuando tratamos de ser explícitos acerca de lo que ella significa, el efecto es mucho más débil? "Hagan eso", dice Barfield, "y casi todo el brillo, y con él la mitad de la poesía, se habrá perdido."[14] ¿Por qué Black piensa que una paráfrasis literal "inevitablemente dice demasiado, y con el énfasis equivocado"? ¿Por qué inevitablemente? ¿Acaso no podemos, si somos lo suficientemente inteligentes, acercarnos tanto como lo deseamos?

A ese respecto, ¿cómo es que el símil se la pasa sin un significado intermedio especial? En general, los críticos no sugieren que el símil dice una cosa y significa otra; ellos no suponen que *signifique* otra cosa que lo que yace en la superficie de las palabras. El símil puede atraernos hacia pensamientos profundos, tal como lo hace una metáfora; ¿cómo es, entonces, que nadie apela al "contenido cognitivo especial" del símil? Recordemos también al hipopótamo de Eliot; allí no había ni símil ni metáfora, pero lo que parecía ocurrir es lo mismo que ocurre cuando hay símiles y metáforas. ¿Acaso alguien sugiere que las *palabras* del poema de Eliot tienen significados especiales?

Por último, si las palabras de la metáfora llevan un significado codificado, ¿cómo puede este significado diferir del significado que esas mismas palabras llevan en el caso en que la metáfora *muere*, esto es, cuando pasa a ser parte del lenguaje? ¿Por qué "El está ardiendo" tal como se lo usa y se lo entiende ahora no significa *exactamente* lo mismo que significaba en una época la metáfora viva? Al fin y al cabo, todo lo que la metáfora muerta significa es que él está muy enojado, una noción no muy difícil de explicitar.

Encontramos así una tensión en la idea usual de metáfora, pues por un lado esta idea pretende sostener que una metáfora hace algo que la prosa común no puede hacer y, por el otro lado, quiere explicar lo que hace una metáfora apelando a un contenido cognitivo: justamente el tipo de cosa que está destinada a expresar la prosa común. Mientras permanezcamos dentro de este marco mental, debemos abrigar la sospecha de ello *puede* hacerse, al menos hasta cierto punto.

Hay una vía sencilla para escapar a este *impasse*. Debemos dejar de lado la idea de que una metáfora transporta un mensaje, de que tiene un contenido o significado (excepto, desde luego, su significado literal). Las diversas teorías que hemos estado considerando confunden su

[13] *Ibídem*, 46.
[14] O. Barfield, "Poetic Diction and Legal Fiction", 55.

meta. Cuando ellas piensan que nos proporcionan un método para descifrar un contenido codificado, en realidad nos dicen (o tratan de decirnos) algo acerca de los *efectos* que las metáforas producen sobre nosotros. El error común consiste en aferrarse a los contenidos de los pensamientos que provoca una metáfora y en leer estos contenidos dentro de la propia metáfora. Sin duda, a menudo las metáforas nos hacen notar aspectos de las cosas que no habíamos notado antes; sin duda atraen nuestra atención hacia analogías y similitudes sorprendentes; efectivamente proporcionan una especie de lente, como dice Black, a través del cual vemos los fenómenos relevantes. El asunto principal no descansa aquí sino en la cuestión de cómo se relaciona la metáfora con lo que ella nos hace ver.

Puede observarse con justicia que la afirmación de que una metáfora provoca o invita a una determinada visión de su sujeto en lugar a decirlo directamente es un lugar común; efectivamente lo es. Así, Aristóteles dice que la metáfora lleva a una "percepción de semejanzas". Black, siguiendo a Richards, dice que una metáfora "evoca" una cierta respuesta: "un oyente apto se verá guiado por una metáfora a construir un...sistema."[15] Este punto de vista se resume nítidamente en lo que Heráclito dijo del oráculo de Delfos: "No dice ni oculta, insinúa."[16]

No pongo en discusión estas descripciones de los efectos de la metáfora, sólo lo hago respecto de las ideas asociadas acerca de *cómo* se supone que la metáfora los produce. Lo que yo niego es que la metáfora haga su trabajo en virtud de tener un significado especial, un contenido cognitivo específico. Yo no pienso, como lo hace Richards, que la metáfora produce su resultado porque tiene un significado que resulta de la interacción de dos ideas; desde mi punto de vista es un error decir como Owen Barfield que una metáfora "dice una cosa y significa otra"; o como Black que una metáfora afirma o implica determinadas cosas complejas gracias a un significado especial y que *por ello* cumple su función de producir una "percepción". Una metáfora hace su trabajo por medio de otros intermediarios; suponer que ella sólo puede ser efectiva si conduce un mensaje codificado es como pensar que una broma o un sueño emiten un enunciado que un intérprete lúcido puede reenunciar en prosa llana. Como una imagen o un golpe en la cabeza, una broma o un sueño o una metáfora pueden hacernos apreciar un hecho, pero sin representar o expresar ese hecho.

Si esto es cierto, lo que intentamos al "parafrasear" una metáfora no puede ser dar su significado, pues eso se encuentra en su superficie; más bien intentamos evocar aquello que la metáfora nos lleva a atender. Puedo imaginar a alguien concediendo esto y desechándolo como si no fuera más que una insistencia en la restricción del uso de la palabra "significado". Esto estaría mal. El error central respecto de la metáfora

[15] M. Black, "Metaphor", 41.

[16] Uso la atractiva traducción que Hannah Arendt da a "σημαίνει"; está claro que no debería traducirse como "significa" en este contexto.

se ataca más fácilmente cuando toma la forma de una teoría del significado metafórico, pero detrás de esa teoría está la tesis, enunciable independientemente, que asociada a una metáfora da forma a un contenido cognitivo definido que su autor desea transmitir y que el intérprete debe captar si quiere tomar el mensaje. Esta teoría es falsa como consideración completa de la metáfora, llamemos o no significado al pretendido contenido cognitivo.

El hecho de que resulte tan difícil decidir exactamente qué se supone que es el contenido, aun en el caso de las metáforas más simples, tendría que hacernos sospechar de la teoría. Pienso que la razón de esta dificultad radica en que nosotros imaginamos que hay un contenido a captar cuando de hecho nos concentramos todo el tiempo en lo que la metáfora nos hace notar. Si lo que la metáfora nos hace notar fuera de alcance finito y de naturaleza proposicional, esto no causaría problemas por sí mismo; simplemente proyectaríamos sobre la metáfora el contenido que la metáfora nos llevó a concebir. Pero en efecto no hay límites para lo que la metáfora atrae a nuestra atención, y la mayor parte de lo que se nos hace notar es de carácter no proposicional. Cuando tratamos de decir qué "significa" una metáfora, enseguida nos damos cuenta de que lo que queremos mencionar no tiene fin.[17] Cuando alguien recorre con su dedo una línea costera en un mapa, o menciona la belleza y la maestría de una línea en un grabado de Picasso, ¿cuántas cosas atraen su atención? Usted podría enumerar una gran cantidad, pero no podría terminar la lista pues la idea de terminar no tendría una aplicación clara. ¿Cuántos hechos o proposiciones transmite una fotografía? ¿Ninguno, infinitos, o un gran hecho indefinible? No es una buena pregunta. Una imagen no vale lo que mil palabras, ni ninguna otra cantidad de ellas. Las palabras no son la moneda apropiada para intercambiar por una imagen.

No sólo se trata de nuestra incapacidad de proporcionar un catálogo exhaustivo de aquello a lo que hemos prestado atención cuando se nos ha hecho ver algo bajo una nueva luz; la dificultad es más fundamental. En general, lo que notamos y vemos no es de caracter proposicional. Por supuesto que *puede* serlo, y cuando lo es, se lo puede enunciar usualmente en palabras bastantes llanas. Pero si yo le muestro a usted el pato-conejo de Wittgenstein, y digo: "Es un pato", entonces con un poco de suerte usted lo verá como un pato; si yo digo: "Es un conejo", usted lo ve como un conejo. Pero ninguna proposición expresa lo que yo

[17] Stanley Cavell menciona el hecho de que la mayor parte de los intentos de abordar la paráfrasis terminan con "y así sucesivamente" y refiere la observación de Empson de que las metáforas están "preñadas" (*Aesthetic Problems of Modern Philosophy*, 79). Pero Cavell no explica la interminabilidad de la paráfrasis como yo lo hago, tal como puede advertirse en el hecho de que él piensa que ella distingue la metáfora de cierto ("pero quizá no todo") discurso literal. Yo sostengo que el caracter interminable de lo que llamamos la paráfrasis de una metáfora surge del hecho de que ella intenta descifrar lo que la metáfora nos hace notar, y para esto no hay un final claro. Yo diría lo mismo para todo uso del lenguaje.

le he llevado a ver. Tal vez usted haya comenzado a darse cuenta de que el dibujo puede verse como un pato o como un conejo. Pero uno podría llegar a saber esto sin haber visto nunca al dibujo como un pato o como un conejo. Ver cómo no es ver qué. La metáfora nos hace ver una cosa como otra haciendo algún enunciado literal que inspira o impulsa la percepción. Dado que en la mayor parte de los casos lo que la metáfora inspira o impulsa no es del todo, incluso no es en absoluto, el reconocimiento de alguna verdad o hecho, el intento de dar expresión literal al contenido de la metáfora está simplemente mal encaminado.

El teórico que trata de explicar una metáfora apelando a un mensaje oculto, al igual que el crítico que trata de enunciar el mensaje, está así fundamentalmente confundido. Ninguna explicación o enunciado de ese tipo puede aparecer en escena pues tal mensaje no existe.

No se trata, claro, de que la interpretación y dilucidación de una metáfora no estén en regla. Muchos de nosotros necesitamos ayuda para ver lo que el autor de una metáfora quiso que viéramos y que un lector más sensible o educado puede captar. La función legítima de la llamada paráfrasis consiste en hacer que el lector perezoso o ignorante tenga una visión semejante a la del crítico instruido. El crítico se encuentra, por así decirlo, en una competencia benigna con el hacedor de la metáfora. El crítico trata de hacer su propio arte, más fácil o más transparente en algunos aspectos que el original, pero al mismo tiempo trata de reproducir en otros algunos de los efectos que el original produjo en él. Al hacer esto, el crítico también orienta la atención, quizá mediante el mejor método que tiene a su alcance, hacia la belleza o la idoneidad, hacia el poder oculto, de la metáfora misma.

18

Comunicación y convención

La convención figura conspicuamente en muchas de nuestras actividades, por ejemplo en la práctica del tarot, en el hablar y en el comer. En la práctica del tarot la convención es esencial, en el comer no lo es. Al explicar cómo se juega al tarot no podemos dejar de tener en cuenta las reglas que definen el juego; al explicar qué es comer no se necesita mencionar reglas o convenciones. ¿Qué sucede en el caso del habla? Las convenciones, ¿son meras comodidades o adornos sociales, o son necesarias para la existencia de la comunicación mediante el lenguaje?

La pregunta es delicada porque no concierne a la verdad de la afirmación de que el habla está sujeta a las convenciones, sino a la importancia y al papel de la convención en el habla. El punto puede presentarse contrafácticamente: ¿podría haber comunicación por medio del lenguaje sin la convención? De acuerdo con David Lewis, "Es una verdad trivial —algo que sólo un filósofo se animaría a negar— decir que hay convenciones de lenguaje".[1] Ciertamente sería absurdo negar que muchas convenciones *involucran* al habla, como el decir "Buenos días" al margen de cómo esté el clima; pero éste no es el tipo de convención del cual depende la existencia del lenguaje. Sin duda, lo que Lewis tiene en mente es la idea que la conexión entre las palabras y lo que ellas significan es convencional. Y quizá sólo un filósofo negaría esto; pero la razón de ello puede ser que sólo un filósofo lo diría por primera vez. Lo que *es* lo suficientemente obvio como para constituir una trivialidad es que el uso de un sonido particular para referir, o significar, lo que refiere o significa es *arbitrario*. Pero en tanto lo convencional es en cierto sentido arbitrario, lo arbitrario no es necesariamente convencional.

En un sentido, describimos completamente un lenguaje cuando decimos qué cuenta como emisión significativa y qué significa cada emisión real o potencial. Pero tales descripciones suponen que nosotros ya conocemos qué representa para una emisión tener un significado particular. La resolución de *esta* cuestión —el problema tradicional del significado— requiere que conectemos la noción de significado con creencias, deseos, intenciones y propósito de un modo que resulte iluminador. Donde más hace falta que la convención desempeñe su tarea es en la concre-

[1] D. Lewis, "Languages and Language", 7.

ción de esta conexión, o conexiones, entre el significado lingüístico y las actitudes y actos humanos descritos en términos no lingüísticos. Y a este respecto se han propuesto muchas teorías diferentes. Las dividiré en tres tipos: primero, están las teorías que afirman que hay una convención que conecta las oraciones de algún modo gramático (o que contiene una frase ejecutora explícita) con intenciones ilocutorias, o con algún propósito más amplio; segundo, están las teorías que buscan un uso convencional para cada oración; y tercero, están las teorías para las cuales hay una convención que vincula las palabras individuales con una extensión o intensión. Estas teorías no compiten entre sí. Todas las combinaciones de estas teorías son posibles, dependiendo de los detalles. Discutiré los tres tipos de teorías en el orden que acabo de enumerar.

En un capítulo temprano y muy influyente, Michael Dummett sostenía que hay una convención que gobierna el uso que damos a las oraciones declarativas.[2] Tal como lo expresa más recientemente:

> ...la emisión de una oración [declarativa] no necesita de un contexto particular para llegar a algo... La emisión de una oración sirve para afirmar algo... hay una convención general por lo cual la emisión de una oración, excepto en contextos especiales, se entiende que es realizada con la intención de emitir una oración verdadera.[3]

Este *dictum* es complejo y tal vez no del todo transparente, pero lo interpretaré como sigue. Existe una conexión convencional entre el emitir una oración declarativa y el usarla para hacer una aserción (uno está haciendo una aserción, excepto en contextos especiales); y hay una conexión conceptual (y tal vez convencional) entre el hacer una aserción y la intención de decir qué es verdadero. La admisibilidad de esta interpretación surge, a mi entender, del más convincente de los argumentos de Dummett. Comienza por examinar las definiciones de verdad al estilo de Tarski. Dummett nos recuerda (siguiendo, si bien problemente no lo sabía, un trabajo previo de Max Black)[4] que si bien Tarski mostró, en principio, cómo construir una definición de verdad para los lenguajes particulares (formalizados), no definió la verdad en general, y además probó que nadie podría hacerlo, al menos usando su método. Por ello Tarski no fue capaz de decir qué era lo que hacía a cada definición de la verdad una definición del mismo concepto. La Convención T, a la cual apeló Tarski como criterio de la corrección de una definición de la verdad, no especifica qué es la verdad en general, pero hace uso de la captación intuitiva que tenemos de ese concepto.

Dummett extrajo una analogía entre la verdad y el concepto de ganar un juego. Si queremos saber qué es ganar un juego, no nos quedaremos satisfechos si nos cuentan la definición de ganar para varios juegos

[2] M. Dummett, "Truth".
[3] M. Dummett, *Frege: Philosophy of Language*, 298.
[4] M. Black, "The Semantic Definition of Truth".

diversos; queremos saber qué hace que la situación definida para cada juego sea un caso de triunfo. Volviendo a la verdad, el problema podría presentarse de esta forma: si nos encontráramos frente a hablantes de un lenguaje que no conociéramos, y se nos diera una definición de la verdad al estilo de Tarski, ¿cómo podríamos determinar si la definición es aplicable a ese lenguaje? Es una buena pregunta; pero yo no creo que se la pueda responder siguiendo la convención propuesta por Dummett, pues a mí me parece que en el lenguaje no hay nada que corresponda de manera relevante a ganar un juego. El punto es importante puesto que si Dummett tiene razón, comprender qué se parece en el lenguaje a ganar un juego equivale a hacer la fundamental conexión entre el significado según lo describe una teoría de la verdad y el uso del lenguaje en los contextos de comunicación.[5]

Ganar en un juego como el ajedrez tiene estas características: primero, la gente que juega usualmente quiere ganar. Quieran o no ganar, es una condición del juego que ellos se *representen* como queriendo ganar. Esto no es lo mismo que fingir que quieren ganar, o que tratar de que los otros crean que ellos quieren ganar. Pero tal vez representarse como queriendo ganar efectivamente implica que uno puede recibir reproches si se descubre que uno no quiere ganar o no está tratando de ganar. Segundo, uno sólo puede ganar si hace movimientos que están definidos por las reglas del juego, y el triunfo está totalmente definido por las reglas. Por último, el triunfo puede ser un fin en sí mismo, y a menudo lo es.[6] Hasta donde yo puedo ver, ninguna conducta lingüística presenta esta combinación de características; de ser así, la analogía de Dummett entre los juegos y el lenguaje es radicalmente deficiente.

Decir la verdad, en el sentido de emitir intencionalmente una oración que resulta ser verdadera, ¿es como ganar? Es en este sentido, que lo que significa decir la verdad es lo que una teoría de la verdad aspira a definir. *Hasta tanto, luego,* los hablantes e intérpretes conozcan de antemano las condiciones de verdad de las emisiones y exista acuerdo sobre ellas en tanto condición para la comunicación, el decir de verdad tiene una de las características del triunfo. (Más adelante cuestionaré hasta qué punto es esto verdadero.) Pero carece de las otras, pues la gente que emite una oración no siempre quiere decir oraciones verdaderas. A veces quiere hacerlo, y muy a menudo no quiere hacerlo. Y para poder jugar al juego del habla tampoco tiene que representarse como si pretendiera o quisiera hablar la verdad; no hay un supuesto general de que alguien que emite una oración declarativa quiere o pretende decir la verdad, ni de que, si lo hace, lo hace intencionalmente. Por último, decir

[5] El hecho de que Dummet no crea que una teoría de la verdad pueda servir como teoría del significado no tiene aplicación en mi argumento. El tema es aquí si hay o no una convención de algún tipo que gobierna nuestras emisiones de oraciones (declarativas).

[6] La distinción entre actividades que pueden ser fines en sí mismas, como tocar la flauta, y aquellas que sirven a otro fin, como construir una casa, proviene, desde luego, de Aristóteles: *Etica Nicomaquea* 1094a; *Magna Moralia* 1211b.

la verdad, en el sentido de emitir una oración verdadera, nunca es un fin en sí mismo.

A diferencia del hablar la verdad, la aserción puede parecer un candidato más apto como contraparte lingüística del triunfo. Alguien que hace una aserción se representa como creyendo lo que dice, y tal vez como justificado en su creencia. Y dado que queremos que nuestras creencias sean verdaderas, parece correcto acordar con Dummett que cuando alguien hace una aserción, se representa como pretendiendo decir lo que es verdadero. (Así interpreto la observación de Dummett de que "se entiende" al hablante con la intención de emitir una oración verdadera.) Igual que cuando se juega un juego, la representación puede o no ser engañosa. (El mentiroso hace una aserción.) El asertor puede pretender o no, cuando hace su aserción, causar que su oyente crea que él cree lo que dice. Así, hacer una aserción es como jugar un juego en un sentido en el cual hablar la verdad no lo es: hay una presunción pública de propósito. En otros aspectos, sin embargo, la aserción no es como ganar, porque lo que constituye la elaboración de una aserción no se encuentra gobernado por reglas o convenciones acordadas.[7]

Para que el concepto de aserción proporcione un puente convencional entre el propósito y la verdad, deben concurrir dos cosas: debe haber convenciones que gobiernen la aserción, y debe haber una convención que vincule la aserción con aquello que se cree verdadero. Yo pienso que ninguna de estas pretensiones está en vigor.

Muchos filósofos han pensado que existen convenciones que gobiernan la aserción. Así, Dummett dice, en una frase que omití en una cita anterior: "La emisión de una oración sirve para afirmar algo..."[8] Consideremos primero si la aserción está gobernada por convenciones. Desde luego, si el que una oración signifique lo que significa literalmente cuando se la emite es una convención, entonces la convención está involucrada en todas las emisiones y por lo tanto también lo está en la aserción. Pero el significado literal puede no ir más allá de las condiciones de verdad (y desde mi punto de vista no lo hace). Y supongo que nadie negará que una misma oración declarativa puede tener el mismo significado cuando se la usa para hacer una aserción, para contar una broma, para fastidiar a un pelmazo, para completar una rima o para formular una pregunta. Por lo tanto, si hay una convención, deben ser otros ornamentos convencionales de la emisión los que la convierten en una aserción. Desde luego, no basta con decir que hay *algo* en el contexto que la vuelve una aserción. Esto es verdad, pero no prueba nada acerca de la convención. Y podemos estar en capacidad de decir qué hay en el contexto que la vuelve una aserción, aunque de hecho yo pienso que sólo podemos decir algunas cosas más bien vagas e incompletas. Pero incluso si las condiciones necesarias y suficientes fueran explícitas y estuvieran acordadas por todas las partes, todavía no se seguiría de ello que las

[7] Con respecto a este punto, mucho debo a Sue Larson.
[8] Véase la nota a pie de página 3.

condiciones son convencionales. Todos estamos de acuerdo en que un caballo debe tener cuatro patas, pero el que los caballos tengan cuatro patas no es una convención.

Hay otra cosa acerca de la aserción que sugiere que puede estar involucrada la convención, cual es el hecho de que al hacer una aserción, el asertor debe proponerse hacer una aserción, y debe proponerse que esta intención sea reconocida por su audiencia. Se pretende que las aserciones sea ejecuciones públicas donde las claves resultan adecuadas para identificar el carácter de la ejecución como asertórico. Por esto es natural pensar que sería útil que existiera una convención, en tanto convendría al propósito de que nuestras intenciones asertivas resulten claras.

Pero con seguridad Frege estaba en lo cierto cuando dijo: "No hay ninguna palabra o signo en el lenguaje cuya función sea simplemente afirmar algo". Frege, como sabemos, decidió más tarde rectificarse e inventó un signo de ese tipo, al molinete " –". Es que entonces Frege estaba trabajando en base a un principio cabal: si hay un rasgo convencional del lenguaje, éste puede hacerse manifiesto en el simbolismo. No obstante, antes de inventar el signo de aserción Frege debió haberse preguntado por qué dicho signo no existía antes. Imaginemos esto: el actor está representando una escena en la cual se supone que hay un incendio. (*Tiny Alice,* de Albee, por ejemplo.) Su papel consiste en imitar con la mayor persuasión posible a un hombre que está tratando de alertar a otras personas sobre el incendio. "¡Fuego!", grita. Y tal vez agrega, a instancias del autor: "¡Lo digo en serio! ¡Vean el humo!", etcétera. Ahora se desata un incendio real, y el actor trata vanamente de alertar a la audiencia real. "¡Fuego!", grita, "¡Lo digo en serio! ¡Vean el humo!", etcétera. Ojalá tuviera el signo de aserción de Frege

Debería resultar obvio que el signo de aserción no sería útil, puesto que el actor lo habría usado en primer lugar, cuando sólo estaba actuando. Un razonamiento similar debería convencernos de que no sería de ayuda decir que el escenario, o el arco del proscenio, crean una disposición convencional que niega la convención de la aserción. Pues si eso fuera así, la convención del teatro también podría expresarse en símbolos; y damos por supuesto que ningún actor o director la usaría. El compromiso del actor está siempre con nosotros. No hay una convención conocida, acordada o públicamente reconocible para hacer aserciones. O, a ese respecto, tampoco las hay para dar órdenes, formular preguntas o hacer promesas. Nosotros hacemos todas estas cosas, a menudo con éxito, y nuestro éxito depende en parte de que hayamos hecho pública nuestra intención de hacerlas. No debemos ese éxito a una convención.

El segundo punto de la declaración de Dummett afirma que hay una convención por la cual se entiende que cuando un hablante hace una aserción está hablando con "la intención de emitir una oración verdadera". Esto también me parece equivocado, aunque de un modo algo diferente. Lo que se entiende es que el hablante, si ha afirmado algo, se ha representado creyéndolo, esto es, emitiendo una oración que cree verdadera. Pero esto no es una convención, sino simplemente parte del

análisis de lo que es la aserción. Afirmar es, entre otras cosas, representarse creyendo lo que uno afirma. Está claro que no puede haber un signo convencional que muestre que uno está diciendo lo que cree; de ser así, todo mentiroso lo usaría. Una convención no puede conectar lo que puede ser siempre secreto —la intención de decir lo que es verdadero— con lo que debe ser público— hacer una aserción—. No hay una convención de la sinceridad.

Si el significado es literal, luego la diferencia entre los modos gramáticos —declarativo, imperativo, interrogativo, optativo— es convencional. Estas diferencias están a la vista y se pretende que sean reconocidas; este trabajo lo hace usualmente la sintaxis sola. Esto evidencia que el modo gramático y la fuerza ilocutoria, por cercana que sea su relación, no pueden relacionarse simplemente por convención.

Si bien he concentrado mis esfuerzos en la aserción, se aplican consideraciones similares a todos los tipos de fuerzas ilocutorias. Sin embargo, mi principal interés no está aquí en la naturaleza de la fuerza ilocutoria o en actos tales como afirmar, prometer y ordenar, sino en la idea de que la convención puede vincular lo que nuestras palabras significan (sus propiedades semánticas literales, incluida la verdad) con nuestros propósitos para usarlas, por ejemplo, para decir la verdad.

Hemos estado discutiendo opiniones que afirman que hay propósitos abarcadores ligados por convención a la empresa de la comunicación lingüística: propósitos que, según palabras de Dummett, nos dan el *quid* del uso del lenguaje. Me vuelco ahora hacia teorías muy distintas, que intentan derivar los significados literales de oraciones enteras (no sólo de los indicadores de modo) de los propósitos no lingüísticos que sirven sus emisiones. En este ensayo me intereso por las teorías que afirman que esa derivación depende de la convención.

Expresadas crudamente, tales teorías sostienen que hay un uso único (o un número finito de usos) al que está vinculada una oración dada, y que ese uso da el significado de la oración. Dado que de hecho pueden darse innumerables usos a una oración, sin modificar su significado, la conexión entre un uso único (o un número finito de usos) y la oración es convencional; a este uso se lo puede llamar estándar.

Esto es demasiado simple, desde luego, pero es una idea atractiva y natural. Pues en efecto parece hacer una importante conexión entre una oración como "Cómete tu berenjena" y la intención, al emitir esta oración, de hacer que alguien coma su berenjena. Hacer que alguien coma su berenjena se corresponde, uno podría decir, al fin para el cual la oración castellana "Cómete tu berenjena" fue concebida. Si esta intuición pudiera enunciarse y defenderse explícitamente sin considerarla presupuesta, sería factible una consideración del significado literal en términos de los propósitos no lingüísticos ordinarios que siempre están detrás de las emisiones de oraciones.

Existen intenciones anidadas en todas las emisiones lingüísticas, tales que si pudiéramos detectarlas usualmente sabríamos qué significan literalmente las palabras emitidas. Pues alguien no puede emitir la oración "Cómete tu berenjena" cuyas palabras signifiquen literalmente

que alguien tiene que comer berenjenas, a menos que pretenda que la oración tenga ese significado, y pretenda que su audiencia interprete que tiene ese significado. Por supuesto que la mera intención no *da* ese significado a la oración; pero si se la admite con la intención de emitir una oración con ese significado, y en efecto no tiene ese significado, entonces no tiene significado lingüístico alguno. El significado literal y el significado literal, pretendido deben coincidir para que haya un significado literal. Pero este hecho, si bien es verdadero e importante, no colabora directamente en la comprensión del concepto de significado literal, puesto que la intención debe estar caracterizada por la referencia al significado literal. Aquí tampoco puede contribuir la convención, pues hemos acudido a ella para convertir los propósitos no lingüísticos en ejecuciones con un significado literal. Una convención que conectara la pretensión de usar palabras con un determinado significado literal con el significado literal de dichas palabras no explicaría el concepto del significado literal, sino que dependería de él.

Lo que estamos buscando son intenciones caracterizadas en términos no lingüísticos: *propósitos ulteriores* en la emisión de oraciones. (Este concepto puede relacionarse con lo que Austin llamó actos perlocutorios.)

Quiero insistir ahora en algo que antes mencioné brevemente; el hecho de que las emisiones lingüística siempre tienen un propósito ulterior; esta fue una de las razones que ofrecí cuando afirmé que ninguna actividad puramente lingüística equivale a ganar un juego. Tal vez esto encierre algún elemento de estipulación, pero yo no hablaría de acto lingüístico si uno dijera "palabras" sólo para oír los sonidos, o para hacer que alguien se durmiera; una acción sólo puede contarse como lingüística si el significado literal es relevante. Pero donde el significado es relevante siempre hay un propósito ulterior. Cuando uno habla, uno aspira a instruir, impresionar, entretener, insultar, persuadir, advertir, recordar, o asistir en un cálculo. Uno incluso puede hablar con la intención de aburrir a una audiencia; pero no con la esperanza de que nadie preste atención al significado.

Si yo estoy en lo cierto cuando digo que cada uso del lenguaje tiene un propósito ulterior, luego uno siempre debe pretender producir algún efecto no lingüístico como resultado de la interpretación de las propias palabras. Max Black lo ha negado, señalando que "...un hombre puede ensayar una conferencia, o escribir una nota para recordar una cita, o simplemente emitir determinadas palabras, como '¡Qué hermoso día!' en ausencia de una audiencia."[9] Los dos primeros casos son sin duda casos en los cuales el significado interesa, y hay una audiencia de la que se pretende que interprete las palabras: uno mismo en un momento posterior. En el último caso resultaría tendencioso insistir en que uno está hablando consigo mismo; sin embargo, interesa qué palabras

[9] M. Black, "Meaning and Intention: An Examination of Grice's Views", 264.

se usan, qué significan. Y debe haber alguna *razón* para usar esas palabras, con sus significados, y en otras. Black cita a Chomsky a este respecto:

> Si bien la consideración de los efectos pretendidos evita algunos problemas, en el mejor de los casos proporcionará un análisis de la comunicación exitosa, pero no del significado o del uso del lenguaje, que no necesitan involucrar a la comunicación, ni siquiera al intento de comunicar. Si yo uso lenguaje para expresar o clarificar mis pensamientos, o con la intención de engañar o para evitar un silencio embarazoso, o en una docena de otras formas, mis palabras tienen un significado estricto y puedo muy bien significar lo que digo, pero la más acabada comprensión de lo que yo pretendo que mi audiencia (si la hay) crea o haga podría ofrecer escasos indicios, o ninguno, del significado de mi discurso.[10]

Me parece que en este pasaje Chomsky arriba a una conclusión correcta a partir de premisas confusas o irrelevantes. El tema es si los significados de las oraciones pueden o no derivarse de las intenciones no linguísticas de un hablante. Chomsky concluye, a mi entender correctamente, que no es posible. Pero ello es irrelevante para la conclusión respecto de si los efectos pretendidos deben involucrar a alguien aparte del hablante, y no es importante para el argumento acerca de la existencia de efectos pretendidos. Hablar o escribir con miras a clarificar pensamientos es ciertamente un supuesto de efecto pretendido. Tampoco importa cómo usemos exactamente la palabra "comunicar". Lo que importa es si una actividad se considera de interés lingüístico cuando no se pretende poner en uso los significados. La mentira representa un caso en que el significado literal es esencial; el mentiroso tiene un propósito ulterior que sólo se satisface si sus palabras se comprenden con el significado que él pretende.

Chomsky tiene razón, como dije, cuando afirma que ninguna cantidad de conocimiento de lo que yo pretendo que mi audiendia crea o haga producirá necesariamente el significado literal de mi emisión. Como hemos visto, incluso esta afirmación debe limitarse a una descripción de mi intención en términos no lingüísticos. Puesto que si yo pretendo que mi audiencia haga o crea algo, ello debe ocurrir a través de su correcta interpretación del significado literal de mis palabras.

Ahora ha quedado relativamente claro qué es lo que debe hacer una convención para relacionar los propósitos no lingüísticos de la emisión de oraciones —los propósitos ulteriores— con los significados literales que esas oraciones tienen cuando se las emite. La convención debe elegir, de un modo comprensible tanto para el hablante como para el oyente y que resulte intencionalmente identificable, aquellos casos en los cuales el propósito ulterior directamente produzca el significado literal. Quiero decir, por ejemplo, un caso donde, el emitir las palabras "Có-

[10] N. Chomsky, *Problems of Knowledge and Freedom*, 19.

mete tu berenjena" con su significado normal en castellano, un hablante pretende hacer que un oyente coma su berenjena mediante la comprensión de las palabras y de la fuerza ilocutoria de la emisión por parte de ese oyente. Y aquí, una vez más, me parece que no sólo no hay tal convención, sino que tampoco puede haberla. Pues incluso si, contrariamente a lo que he argumentado, alguna convención gobernara la fuerza ilocutoria de la emisión, la conexión con la intención de que la petición o la orden fuera llevada a cabo requeriría que el hablante fuera *sincero*: que aquello que él se representa como queriendo o tratando de hacer en efecto quisiera o tratara de hacerlo. Pero no hay nada más obvio que la imposibilidad de que haya una convención que señale la sinceridad.

Debo repetir que no nos ayuda nada decir que la convención consiste en que la oración siempre significa lo que el propósito ulterior revelaría si el hablante fuera sincero, serio, etcétera. Esto es a lo sumo un *análisis* parcial de las relaciones entre significado literal, sinceridad e intención. No sugiere pruebas, criterios o prácticas públicamente reconocidos.

A veces se sugiere que un lenguaje sólo puede aprenderse en una atmósfera de aserciones honestas (órdenes, promesas, etcétera). Aunque esto fuera verdad, nada probaría acerca de un supuesto rol de la convención. Pero además me declaro escéptico frente a esta pretensión, en parte porque una gran proporción del aprendizaje del lenguaje tiene lugar en los juegos, cuando se escuchan cuentos y en la ficción, y en parte porque la adquisición del lenguaje no puede depender hasta ese punto de nuestra suerte de tener compañeros de juego y padres veraces, sobrios y asertivos.

La naturaleza de un juego como el ajedrez o el tarot incluye no sólo que haya criterios mutuamente acordados de lo que es jugar, sino también criterios mutuamente acordados de lo que es ganar. Para estos juegos es esencial que normalmente no haya cuestionamientos acerca de los resultados. Y también forma parte de su naturaleza el que la victoria pueda ser un fin en sí mismo, y que los jugadores se representen como queriendo o tratando de ganar. Pero los criterios para decidir lo que una emisión significa literalmente, que una teoría de la verdad o del significado da a un hablante, no deciden si éste ha alcanzado su propósito ulterior, así como tampoco hay una regla general por la cual los hablantes se representen como si tuvieran otro fin más allá de usar las palabras con un determinado significado y fuerza. El propósito ulterior puede o no ser evidente, y puede o no ayudar a un intérprete a determinar el significado literal. Concluyo que no puede ser un rasgo accidental del lenguaje que el propósito ulterior de una emisión y su significado literal sean independientes, en el sentido de que el último no puede derivarse del primero: forma parte de la esencia del lenguaje. Daré a este rasgo del lenguaje el nombre de principio de *autonomía de significado*. Encontramos una aplicación del mismo cuando hablamos de la fuerza ilocutoria, donde tomó la forma del descubrimiento de que lo que se incluye dentro del significado literal queda luego disponible para cual-

quier propósito ulterior (no lingüístico), e incluso para cualquier ejecución ilocutoria.[11]

Antes de terminar la discusión de las teorías de los dos primeros tipos, la siguiente observación puede ser de utilidad. Nada de lo que he dicho ha pretendido mostrar que no existe conexión entre los indicadores de modo y la *idea* de un determinado acto ilocutorio. Yo creo que existe tal conexión. Una emisión de una oración imperativa, por ejemplo, se rotula a sí misma, en forma bastante literal, como un acto de ordenar. Pero esto es sólo una parte del significado literal de las palabras emitidas, y no establece ninguna relación, convencional o de otro tipo, entre las intenciones ilocutorias del hablante y sus palabras. Es fácil confundir dos tesis bastante diferentes: por un lado, la tesis (correcta) de que cada emisión de una imperativa *se rotula* a sí misma (verdadera o falsamente) como una orden, y la tesis de que hay una convención por la cual bajo condiciones "estándar" la emisión de una imperativa *es* una orden. La primera tesis efectivamente explica la diferencia de significado entre una oración imperativa y una declarativa, diferencia que existe de manera totalmente independiente de la fuerza ilocutoria. La segunda tesis no puede explicar esto pues ella postula una convención que rige sólo bajo condiciones "estándar". Uno no puede *usar* una convención violándola, uno sólo puede *abusar* de ella. Pero la diferencia entre una imperativa y una declarativa puede usarse, y muchas veces se lo hace, muy apropiadamente en situaciones donde el modo y la fuerza ilocutoria no son "estándar". No aportará nada señalar casos como el de un actor que lleva una corona para indicar que está haciendo el papel del rey. Si esto involucra una convención, se trata de una convención que gobierna el significado literal. Llevar una corona, sea en broma o en serio, es lo mismo que decir: "Yo soy el rey".

La sustancia de esta observación es aplicable también al tipo de teoría que trata de derivar el significado literal de cada oración de un uso "estándar". Puesto que el significado literal opera tanto cuando el uso está ausente como cuando está presente, ninguna convención que opere solamente en situaciones "estándar" puede darlo.

Hemos considerado la idea de que la actividad lingüística en general es como un juego en cuanto hay un propósito convencional (decir qué es verdadero, ganar) que sólo puede lograrse si se siguen, o se usan, reglas acordadas y públicas. Luego discutimos la pretensión de que el significado literal de cada oración se relaciona mediante una convención a un fin no lingüístico estándar (un propósito ulterior). Tras el examen, ambos puntos de vista resultaron ser insostenibles. Ahora es tiempo de evaluar la "trivialidad" de que el significado de una palabra es convencional, esto es, que asignar el significado que asignamos a las palabras individuales y a las oraciones cuando se las emite o escribe es una convención.

[11] Para una mayor discusión, y una sugerencia de la forma en que se relacionan la fuerza ilocutoria y el modo gramático, véase el Ensayo 8.

De acuerdo con David Lewis[12] una convención es una *regularidad* R en la acción, o en la acción y la creencia, regularidad en la cual debe estar involucrada más de una persona. La regularidad tiene estas propiedades:

(1) Cada persona involucrada se ajusta a R y (2) cree que los otros también se ajustan. (3) La creencia de que los otros se ajustan a R da a todos los involucrados una buena razón para ajustarse a R. (4) Todos los involucrados prefieren que haya ajuste respecto de R. (5) R no es la única regularidad posible que cumple las dos últimas condiciones. (6) Por último, cada persona involucrada conoce (1)-(5) y sabe que todos los demás conocen (1)-(5), etcétera.

Tyler Burge ha manifestado dudas razonables respecto de partes de la última condición (¿una convención requiere que todos sepan que hay alternativas?)[13] y yo siento mis propios recelos. Pero aquí no hay razón para debatir los detalles del análisis del concepto de convención; lo relevante es plantear la cuestión de si la convención nos ayuda en algún sentido razonable a comprender la comunicación lingüística, o en qué grado lo hace. Así, en lugar de preguntar, por ejemplo, que cosa "ajustada" a la regularidad se agrega a la regularidad misma, simplemente concederé que algo como las seis condiciones de Lewis efectivamente vale en general para lo que llamamos hablantes del mismo lenguaje. ¿Cuán fundamental es este hecho para el lenguaje?

El análisis de Lewis requiere claramente que haya por lo menos dos personas involucradas, dado que la convención depende de una práctica mutuamente comprendida. Pero nada en el análisis requiere más de dos personas. Dos personas podrían tener convenciones, y podrían compartir un lenguaje.

¿Qué es exactamente la convención necesaria? No puede ser que el hablante y el oyente signifiquen la misma cosa emitiendo las mismas oraciones. Pues tal ajuste, aunque tal vez sea bastante común, no es necesario para la comunicación. Cada hablante puede hablar un lenguaje diferente, y esto no deteriorará la comunicación en tanto cada oyente comprenda al que habla. Hasta podría suceder que cada hablante tuviera desde un principio su forma propia y singular de hablar. Efectivamente, en la práctica sucede algo parecido a esto. Los diferentes hablantes tienen diferentes repertorios de nombres propios, diferentes vocabularios, y adjudican significados algo diferentes a las palabras. En algunos casos esto reduce el nivel de comprensión mutua; pero no necesariamente, pues como intérpretes somos muy buenos para alcanzar una interpretación correcta de palabras que no hemos escuchado antes, o de palabras que no hemos escuchado antes con los significados que un hablante les está dando.

[12] D. Lewis, "Languages and Language", 5, 6.
[13] T. Burge, "Reasoning about Reasoning".

La comunicación no requiere, entonces, que el hablante y el oyente signifiquen la misma cosa con las mismas palabras; mas la convención requiere un ajuste por parte de al menos dos personas. Queda sin embargo otra forma de acuerdo que resulta necesaria: si la comunicación tiene éxito, el hablante y el oyente deben asignar el mismo significado a las palabras del hablante. Además, como hemos visto, el hablante debe pretender que el oyente interprete sus palabras en la forma que él pretende, y debe tener una razón adecuada para creer que el oyente logrará interpretarlo como él pretende. Tanto el hablante como el oyente deben creer que el hablante habla con esta pretensión, y así sucesivamente; en breve, parecería que muchas de las condiciones de Lewis están satisfechas. Es cierto que esto es a lo sumo una forma atenuada de práctica o convención, alejada de la idea usual de práctica común. Pero todavía uno podría insistir en que este grado de método de interpretación mutuamente comprendido es el núcleo convencional esencial de la comunicación lingüística.

Pero la característica más importante del análisis de la convención de Lewis, la regularidad, aún sigue sin explicarse. La regularidad en este contexto debe significar en el tiempo, y no un mero acuerdo en un momento. Para que haya una convención en el sentido de Lewis (o en cualquier sentido, diría yo), debe verse que algo se repite en el tiempo. El único candidato a la recurrencia que tenemos son los patrones de sonido: el hablante y el oyente deben interpretar repetidamente, intencionalmente, y bajo mutuo acuerdo, patrones de sonido relevantemente similares del hablante de la misma manera (o de maneras relacionadas mediante reglas que puedan hacerse explícitas de antemano).

Yo no dudo de que toda comunicación lingüística humana efectivamente muestra algún grado de esta regularidad, y quizás algunos se sientan inclinados a estipular como condición para que una actividad se considere lingüística la existencia de tal regularidad. Pero tengo mis dudas, tanto respecto de la claridad de la idea como respecto de su importancia en la explicación y descripción de la comunicación. La claridad es cuestionable debido a la gran dificultad de decir exactamente cómo deben coincidir las teorías del hablante y del oyente para la interpretación de las palabras de hablante. Desde luego, ellas deben coincidir *después* de que se ha hecho una emisión, pues de no ser así la comunicación se resiente. Pero a menos que coincidan de antemano, los conceptos de regularidad y convención no representarán una ventaja definida. Con todo, se podrá llegar con seguridad a un acuerdo respecto de lo que un hablante significa con lo que dice, aunque hablante y oyente tengan de antemano diferentes teorías para interpretar al hablante. Esto puede ser así en razón de que el hablante bien puede proporcionar claves adecuadas en lo que dice, cómo y dónde lo dice, que permitan al oyente llegar a una interpretación correcta. Claro que el hablante debe tener *alguna* idea de la forma en que el oyente puede dar uso a las claves relevantes; y el oyente debe conocer bastante bien lo que puede esperar. Pero ese conocimiento general es difícil de reducir a reglas, por no hablar de convenciones o prácticas.

Es fácil juzgar erróneamente el papel de la sociedad en el lenguaje. El lenguaje es, sin duda, un arte social. Pero nos equivocamos si al advertir la forma en que la sociedad convierte hábitos lingüísticos en normas públicas suponemos que hemos penetrado profundamente en el corazón de la comunicación lingüística. Lo convencional en el lenguaje, si algo lo es, es que la gente tiende marcadamente a hablar como lo hacen sus vecinos. Pero al indicar este elemento de lo convencional, o del proceso condicionante que hace de los hablantes simples facsímiles lingüísticos de sus amigos y parientes, no explicamos más que la convergencia; no esclarecemos en nada la naturaleza esencial de las habilidades que se hacen converger de ese modo.

Esto no es negar la importancia práctica del condicionamiento social, a diferencia de su importante teórica. El condicionamiento común garantiza que podemos suponer, hasta cierto punto, que el mismo método de interpretación que usamos para otros, o que suponemos que los otros usan para nosotros, funcionará para un nuevo hablante. No tenemos tiempo, paciencia u oportunidad para desarrollar una nueva teoría de la interpretación para cada hablante, y salimos del paso porque desde el momento en que alguien que nos es desconocido abre la boca, ya tenemos un conocimiento enorme del tipo de teoría que funcionará para él, o sabemos que no conocemos tal teoría. Pero si sus primeras palabras son, por ejemplo, castellanas, podemos suponer justificadamente que ha estado expuesto a un condicionamiento lingüístico similar al nuestro (hasta podemos adivinar o conocer diferencias). Al comprar una pipa, ordenar un plato o dar indicaciones al conductor de un taxi, partimos de esta suposición. Así, hasta que se pruebe su falsedad; en cuyo caso podemos revisar nuestra teoría de lo que él significa actuando según nos dicte el momento. Cuanto más tiempo dure la conversación, mejor resultará nuestra teoría, y más detalladamente se adaptará al hablante individual. Vemos así que el conocimiento de las convenciones del lenguaje constituye un punto de apoyo práctico para la interpretación, sin el cual no podemos pasárnoslas en la práctica; pero que, bajo condiciones óptimas para la comunicación, podemos finalmente eliminar, y en teoría podríamos haber eliminado desde un principio.

El hecho de que la interpretación radical sea hasta tal punto un lugar común —esto es, el hecho de que usemos nuestro método usual de interpretación sólo como un punto de partida útil en la comprensión de un hablante se nos oculta por medio de varios factores, siendo el primer de ellos que la sintaxis es muchísimo más social que la semántica. En términos generales, esto se debe a que aquello que forma el esqueleto de lo que llamamos lenguaje es el patrón de inferencia y estructura creado por las constantes lógicas: conectivos oracionales, cuantificadores y recursos para referencia cruzada. Si podemos aplicar nuestro método general de interpretación siquiera a un hablante, si podemos al menos empezar a comprenderlo en la suposición de que su lenguaje es como el nuestro, esto sucederá porque podemos tratar sus recursos de formación de estructura tal como tratamos los nuestros. Esto fija la forma lógica de sus oraciones y determina las partes del habla. Sin duda, cierto re-

pertorio de predicados importantes debe traducirse en la forma homofónica obvia para que podemos avanzar rápidamente; pero luego podemos arreglárnoslas muy bien intepretando o reinterpretando otros predicados nuevos o aparentemente familiares.

Esta ilustración de la forma en que tiene lugar la interpretación presenta la aplicación de los métodos formales al lenguaje natural bajo otra luz. Ayuda a mostrar por qué los métodos formales encuentran su mejor aplicación en la sintaxis; aquí al menos hay una buena razón para esperar que un mismo modelo se ajuste bastante bien a una cantidad de hablantes. Y no hay una razón clara por la cual cada método hipotético de interpretación no podría ser una semántica formal para lo que en términos generales consideramos un lenguaje. Lo que no podemos pretender, sin embargo, es poder formalizar las consideraciones que nos llevaron a ajustar nuestra teoría para adecuar el influjo de información nueva. No cabe duda de que normalmente consideramos la capacidad de cambiar apropiadamente de terreno como parte de lo que entendemos por "conocer el lenguaje". Pero, en este sentido, no se puede decir qué tiene que conocer alguien que conoce el lenguaje, puesto que la intuición, la suerte y la destreza deben desempeñar aquí un papel tan esencial como el que desempeñan en la concreción de una nueva teoría en cualquier campo, así como el gusto y la simpatía deben desempeñar un papel aún mayor.

A modo de conclusión, deseo refirmar entonces que la comunicación lingüística no requiere de una repetición gobernada por reglas, aunque muy a menudo hace uso de ella; y en ese caso, la convención no ayuda a explicar qué resulta básico para la comunicación lingüística, si bien puede describir un rasgo usual aunque contingente.

Haré una reflexión final. No he hablado aquí —aunque lo he hecho en otras oportunidades— de que no podemos confiarnos en atribuir creencias, deseos e intenciones a una criatura que no puede usar el lenguaje.[14] Las creencias, los deseos y las intenciones son una condición del lenguaje, pero el lenguaje es también una condición para ellos. Por otra parte, ser capaces de atribuir creencias y deseos a una criatura es sin duda una condición que implica compartir una convención con esa criatura; en tanto que, si lo que he dicho en este trabajo es correcto, la convención no es una condición del lenguaje. Sugiero entonces que los filósofos que hacen de la convención un elemento necesario han invertido los papeles. La verdad indica más bien que el lenguaje es una condición para tener convenciones.

[14] Véase el Ensayo 11.

Referencias bibliográficas

Allen, Woody, "Condemned", *New Yorker* (21 de noviembre de 1977), 59.

Arendt, H., *The Life of the Mind,* Harcourt, Brace, Jovanovich, Nueva York, 1978.

Austin, J. L., y Strawson, P. F., "Symposium on 'Truth'", en *Proceedings of the Aristotelian Society,* Volumen Suplementario 24, 1950.

Austin, J. L., *How to Do Things with Words,* Harvard University Press, Cambridge, Massachusetts, 1962.

Bain, A., *English Composition and Rhetoric,* D. Appleton, Nueva York, 1867.

Bar-Hillel, Y., "Logical Syntax and Semantics", *Language,* 30, 1954, 230-7.

Bar-Hillel, Y., "Remarks on Carnap's Logical Syntax of Language", en *The Philosophy of Rudolf Carnap,* comp. P. A. Schilpp, Open Court, La Salle, Illinois, 1963.

Bar-Hillel, Y., *Language and Information,* Addison-Wesley, Jerusalén, 1964.

Barfield, O., "Poetic Diction and Legal Fiction", en *The Importance of Language,* comp. M. Black, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1962.

Benacerraf, P., "Mathematical Truth", *Journal of Philosophy,* 70, 1973, 661-79.

Beth, E. W., "Carnap's Views on the Advantages of Constructed Systems Over Natural Languages in the Philosophy of Science", en *The Philosophy of Rudolf Carnap,* Comp. P. A. Schilpp, Open Court, La Salle, Illinois, 1963.

Black, M., "The Semantic Definition of Truth", *Analysis,* 8, 1948, 49-63.

Black, M., "Metaphor", en *Models and Metaphors,* Cornell University Press, Ithaca, Nueva York, 1962.

Black, M., "Meaning and Intention: an Examination of Grice's Views", en *New Literaty History,* 4, 1972-3, 257-79

Black, M., "How Metaphors Work: A Reply to Donald Davidson", *Critical Inquiry,* 6, 1979, 131-43.

Bohnert, H., "The Semiotic Status of Commands", *Philosophy of Science,* 12, 1945, 302-15.

Burge, T., "Reference and Proper Names", *Journal of Philosophy,* 70, 1973, 425-39.

Burge, T., "Reasoning about Reasoning", *Philosophia,* 8, 1979, 651-6.

Carnap, R., *The Logical Syntax of Language,* Routledge and Kegan Paul, Londres, 1937.

Carnap, R., *Meaning and Necessity,* University of Chicago Press, Chicago, primera edición, 1947, edición aumentada, 1956.

Cartwright, R., "Propositions", en *Analytical Philosophy,* comp. R. J. Butler, Blackwell, Oxford, 1962.

Cavell, S., "Aesthetic Problems of Modern Philosophy", en *Must We Mean What We Say?,* Charles Scribner, Nueva York, 1969.

Cohen, T., "Figurative Speech and Figurative Acts", *Journal of Philosophy,* 72, 1975, 669-84.
Chomsky, N., *Aspects of the Theory of Syntax,* M. I. T. Press, Cambridge, Massachusetts, 1965.
Chomsky, N., "Topics in the Theory of Generative Grammar", en *Current Trends in Linguistics, 3,* comp. T. A. Sebeok, The Hague, Mouton, 1966.
Chomsky, N., *Problems of Knowledge and Freedom,* Pantheon, Nueva York, 1971.
Church, A., "On Carnap's Analysis of Statements of Assertion and Belief", *Analysis,* 10, 1950, 97-9.
Church, A., "A Formulation of the Logic of Sense and Denotation", en *Structure, Method and Meaning: Essays in Honour of H, M. Sheffer,* comp. P. Henle, H, M. Kallen y S. K. Langer, Liberal Arts Press, Nueva York, 1951.
Chuch, A., "Intensional Isomorphism and Identity of Belief", *Philosophical Studies,* 5, 1954, 65-73.
Church, A., *Introduction to Mathematical Logic,* Vol. 1, Princeton University Press, Princeton, 1956.
Davidson, D., *Essays on Actions and Events,* Clarendon Press, Oxford, 1980.
Davidson, D., "Toward a Unified Theory of Meaning and Action", en *Grazer Philosophische Studien,* 2, 1980, 1-12.
Dumett, M., *Frege: Philosophy of Language,* Duckworth, Londres, 1973.
Dummett, M., "Truth", en *Truth and Other Enigmas,* Duckworth, Londres, 1978.
Eliot, T. S., *Selected Poems,* Harcourt, Brace, Jovanovich, Nueva York, 1967.
Empson, W., *Some Versions of Pastoral,* Chatto and Windus, Londres, 1935.
Feyerabend, P., "Explanation, Reduction, and Empiricism", en *Scientific Explanation, Space and Time: Minnesota Studies in the Philosophy of Science, 3,* University of Minnesota Press, Minneapolis, 1962.
Feyerabend, P., "Problems of Empiricism", en *Beyond the Edge of Certainty,* comp. R. G. Colodny, Prentice-Hall, Englewood Cliffs, Nueva Jersey, 1965.
Field, H., "Tarski's Theory of Truth, *Journal of Philosophy,* 69, 1972, 347-75.
Field, H., "Quine and the Correspondence Theory", *Philosophical Review,* 83, 1974, 200-28.
Field, H., "Conventionalism and Instrumentalism in Semantics", *Noûs,* 9, 1975, 375-406.
Foster, J., "Meaning and Truth Theory", en *Truth and Meaning,* comp. G. Evans y J. McDowell, Clarendon Press, Oxford, 1976.
Frege, G., "On sense and Reference", en *Philosophical Writings,* comp. M. Black y P. T. Geach, Blackwell, Oxford, 1962, 56-78.
Geach, P. T., *Mental Acts,* Routledge and Kegan Paul, Londres, 1957.
Geach, P. T., "Assertion", *Philosophical Review,* 74, 1965, 449-65.
Geach, P. T., "Quotation and Quantification", en *Logic Matters,* Blackwell, Oxford, 1972.
Goodman, N., *Languages of Art,* Bobbs-Merrill, Indianapolis, 1968.
Goodman, N., "Metaphor as Moonlighting", *Critical Inquiry,* 6, 1979, 125-30.
Harman, G., "Logical Form", en *Foundations of Language,* 9, 1972, 38-65.
Harman, G., "Meaning and Semantics", en *Semantic and Philosophy,* comps. M. I. Munitz y P. K. Unger, New York University Press, Nueva York, 1974.
Harman, G., "Moral Relativism Defended", *Philosophical Review,* 84, 1975, 3-22.
Henle, P., "Metaphor", en *Language, Thought and Culture,* comp. Henle, University of Michigan Press, Ann Arbor, Michigan, 1958.
Hintikka, J., *Knowledge and Belief,* Cornell University Press, Ithaca, 1962.

Jackson, H., *The Eighteen-Nineties*, Alfred Knopf, Nueva York, 1922.
Jeffrey, R., reseña de Logic, *Methodology, and the Philosophy of Science*, comp. E. Nagel, P. Suppes y A.Tarski, *Journal of Philosophy*, 61, 1964, 79-88.
Jeffrey, R., *The Logic of Decision*, McGraw-Hill, Nueva York, 1965.
Kripke, S., "Is There a Problem about Substitutional Quantification?", en *Truth and Meaning*, comps. G. Evans y J. McDowell, Clarendon Press, Oxford, 1976.
Kuhn, T. S., *The Structure of Scientificic Revolutions*, University of Chicago Press, Chicago, 1962.
Kuhn, T. S., "Reflections on my Critics", en *Criticism and the Growth of Knowledge*, comps. I. Lakatos y A. Musgrave, Cambrige University Press, Cambridge, Inglaterra, 1970.
Lewis, D., "General Semantics", en *Semantics for Natural Language*, comps. D. Davidson y G. Harman, D. Reidel, Dordrecht-Holland, 1972.
Lewis, D., "Radical Interpretation", *Synthèse*, 27, 1974, 331-44.
Lewis, D., "Languages and Language", en *Language, Mind and Knowledge*, comp. K. Gunderson, University of Minnesota Press, Minneapolis, 1975.
McCarrell, N. S., y Verbrugge, R. R., "Metaphoric Comprehension: Studies in Reminding and Resembling", *Cognitive Psychology*, 9, 1977, 494-533.
Malcolm, N., "Thoughtless Brutes", en *Proceedings and Addresses of the American Philosophical association*, 1972-3
Mates, B., "Synonymity", en *Semantics and the Philosophy of Language*, comp. L. Linsky, University of Illinois Press, Urbana, Illinois, 1952.
Mates, B., *Elementary Logic*, Clarendon Press, Oxford, 1965.
Murray, J. Middleton, *Countries of the Mind*, Collins, Londres, 1922.
Onions, C.T., *An Advanced English Syntax*, Routledge and Kegan Paul, Londres, 1965.
Oxford English Dictionary, The, comp. J. A. H. Murray *et al*, Clarendon Press, Oxford, 1933.
Parsons, K. P., "Ambiguity and the Theory of Truth", *Noûs*, 7, 1973, 379-94.
Prior, A. N., *Past, Present and Future*, Clarendon Press, Oxford, 1967.
Putnam, H., "The Meaning of 'Meaning'", en *Mind, Language and Reality*, Cambridge University Press, Cambridge, Inglaterra, 1975.
Quine, W. V., *Mathematical Logic*, Harvard University Press, Cambridge, Massachusetts, 1940.
Quine, W. V., *Methods of Logic*, Holt, Nueva York, 1950.
Quine, W. V., *Word and Object*, M. I. T. Press, Cambridge, Massachusetts, 1960.
Quine, W. V., *From a Logical Point of View*, Segunda edición, Harvard University Press, Cambridge, Massachusetts, 1961.
Quine, W. V., "Two Dogmas of Empiricism", en *From a Logical Point of View*, Segunda edición, Harvard University Press, Cambridge, Massachusetts, 1961.
Quine, W. V., "On an Application of Tarski's Theory of Truth", en *Selected Logic Papers*, Random House, Nueva York, 1966.
Quine, W. V., "Truth by Convention", en *The Ways of Paradox*, Random House, Nueva York, 1966,70-99.
Quine, W. V., "Ontological Relativity", en *Ontological Relativity and Other Essays*, Columbia University Press, Nueva York, 1969.
Quine, W. V., "Reply to 'On Saying That'", en *Words and Objetions*, comps. D. Davidson y G. Harman, D. Reidel, Dordrecht-Holland, 1969.
Quine, W. V., "Speaking of Objects", en *Ontological Relativity and Other Essays*, Columbia University Press, Nueva York, 1969.
Quine, W.V., "Comments on 'Belief and the Basis of Meaning'", *Synthèse*, 27,

1974, 325-9.
Quine, W. V., *The Roots of Reference,* Open Court, La Salle, Illinois, 1974.
Quine, W. V., "On Empirically Equivalent Systems of the World", *Erkenntnis,* 9, 1975, 313-28.
Quine, W. V, "Comments on "Moods and Performances'", en *Meaning and Use,* comp. A. Margalit, D. Reidel, Dordrecht-Holland, 1979.
Quine, W. V., "On the Very Idea of a Third Dogma", en *Theories and Things,* Harvard University Press, Cambridge, Massachusetts, 1981.
Quine, W. V., "Replies to the Eleven Essays", *Southwestern Journal of Philosophy,* 10, 1981.
Ramsey, F. P., "Facts and Propositions", reimpreso en *Foundations of Mathematics,* Humanities Press, Nueva York, 1950.
Ramsey, F. P., "Truth and Probability", reimpreso en *Foundations of Mathematics,* Humanities Press, Nueva York, 1950.
Reichenbach, H., *Elements of Symbolic Logic,* Macmillan, Londres, 1947.
Ross, J. R., "Metalinguistic Anaphora", *Linguistic Inquiry,* 1, 1970, 273.
Scheffler, I., "An Inscriptional Approach to Indirect Quotation", *Analysis,* 10, 1954, 83-90.
Scheffler, I., *The Anatomy of Inquiry,* Alfred Knopf, Nueva York, 1963.
Sellars, W., "Putnam on Synonymity and Belief", *Analysis,* 15, 1955, 117-20.
Sellars, W., "Truth and 'Correspondence'", *Journal of Philosophy,* 59, 1962, 29-56.
Sellars, W., "Conceptual Change", en *Conceptual Change,* comps. G. Pearce y P. Maynard, D. Reidel, Dordrecht-Holland, 1973.
Strawson, P. F., y Austin, J. L., "Symposium on 'Truth'", en *Proceedings of the Aristotelian Society,* Volumen suplementario 24, 1950.
Strawson, P. F., "Singular Terms, Ontology and Identity", *Mind,* 65, 1956, 433-54.
Strawson, P. F., *Individuals,* Methuen, Londres, 1959.
Strawson, P. F., "Truth: A Reconsideration of Austin's Views", *Philosophical Quarterly,* 15, 1965, 289-301.
Strawson, P. F., *The Bounds of Sense,* Methuen, Londres, 1966.
Tarski, A., "The Semantic Conception of Truth", *Philosophy and Phenomenological Research,* 4, 1944, 341-75.
Tarski, A., "The Concept of Truth in Formalized Languages", en *Logic, Semantics, Metamathemics,* Clarendon Press, Oxford, 1956.
Tarski, A., "Truth and Proof", *Scientific American,* 220, 1967, 63-77.
Tharp, L., "Truth, Quantification, and Abstract Objects", *Noûs,* 5, 1971, 363-72.
Thomson, J. F., "Truth-bearers and the Trouble about Propositions", *Journal of Philosophy,* 66, 1969, 737-47.
Verbrugge, R. R, y McCarrell, N. S., "Metaphoric Comprehension: Studies in Reminding and Resembling", *Cognitive Psychology,* 9, 1977, 494-533.
Wallace, J., *Philosophical Grammar,* Ph. D. Thesis, Stanford University, 1964.
Wallace, J., "Propositional Attitudes and Identity", *Journal of Philosophy,* 66, 1969, 145-52.
Wallace, J., "On the Frame of Reference", *Synthèse,* 22, 1970, 61-94.
Wallace, J., "Convention T and Substitutional Quantification", *Noûs,* 5, 1971, 199-21.
Wallace, J., "Positive, Comparative, Superlative", *Journal of Philosophy,* 69, 1972, 773-82.
Wallace, J., "Nonstandard Theories of Truth", en *The Logic of Grammar,* comps. D. Davidson y G. Harman, Dickenson Publishing Co., Belmont, California, 1975.

Wallace, J., "Only in the Context of a Sentence do Word Have Any Meaning" *Midwest Studies in Philosophy, 2: Studies in the Philosophy of Language,* comps. P. A. French, T. E. Uehling, Jr., y H. K. Wettstein, University of Minnesota Press, Morris, 1977.

Weinstein, S., "Truth and Demonstratives", *Noûs,* 8, 1974, 179-84.

Whorf, B. L., "The Punctual and Segmentative Aspects of Verbs in Hopi", en *Language, Thought and Reality: Selected Writing of Benjamin Lee Whorf,* comp. J. B. Carroll, The Techonology Press of Massachusetts Institute of Techonology, Cambridge, Massachusetts, 1956.

Wilson, N. L., "Substances Without Substrata", *Review of Metaphysics,* 12, 1959, 521-39.

Indice temático

Acción: explicación de la, 167-169; y significado, 21, 138; *véase también* teoría de la decisión
Actitudes proposicionales: y animales, 21, 164-178, 276; *véase también* significado y creencia; creencia
Afirmación, *véase* considerar verdadero
ALBEE, E., 267
ALLEN, W., 256
Ambigüedad, 51, 76, 144, 248-249
ARENDT, H., 260n.
ARISTÓTELES, 204. 245, 260, 265
Aserción, 124-134, 173, 206; distinguida del modo indicativo, 124-128; forma lógica de oraciones acerca de la, 121, 131; *véase también* oraciones indicativas
AUSTIN, J. L., 55n., 57n., 72, 124n., 269
Autonomía del significado, *véase* significado, autonomía del

BAIN, A., 252n.
BARFIELD, O., 259, 260
BAR-HILLEL, Y., 49n., 76, 123
BEARDSLEY, M., 246
BENACERRAF, P., 222
BERGSON, H., 190
BERKELEY, G., 223
BETH, E., 49n.
BILGRAMI, A., 23
BLACK, M., 15, 244-245, 252n., 253, 258-260, 269-270
BOHNERT, H., 128
BURGE, T., 143, 273

Caridad, principio de, 21, 48, 115, 147-148, 161-162, 176, 201-202, 204
CARNAP, R., 35, 52, 112, 114, 204, 210n.
CARTWRIGHT, R., 63n.
Causa: teorías causales de la referencia, 236-238; oraciones acerca de la, 215-217
CAVELL, S., 261
Cita, 32-34, 211; como una teoría del discurso indirecto, 112-113; teoría de ilustración de la, 99-100; teoría del nombre propio de la, 97-98; teoría del deletreo de la, 101-104; teoría demostrativa de la, 104-107
COHEN, T., 256
Conductismo, 162, 223, 232
Considerar verdadero, 146, 161, 169-170, 175-176, 201, 232-233
Contextos intensionales, 17, 100, 108-120, 132, 172, 182, 183, 211-212; *véase también* creencia, oraciones acerca de la; discurso indirecto
Convención, 23, 125-128, 263-276; como conexión entre modo y fuerza, 264-268; como conexión entre uso y significado, 269-271; como fijación del significado, 273-276; lenguaje en tanto gobernado por la, 23, 125-128, 263-276; *véase también* aserción; imperativas; interrogativas
Convención T., 17-18, 45, 65, 74, 79, 83-91, 143, 147, 149, 159, 162, 180-182, 200, 209; *véase también* Tarski; teorías de la verdad al estilo de Tarski; verdad
Creencia, 205; en relación al significado, 151-158, 200-201, 239-241; oraciones acerca de la, 36, 51-52, 173-175, 211-212; *véase también* actitudes proposicionales
CHOMSKY, N., 19, 43, 51, 78, 80-81, 269-270
CHURCH, A., 17, 33, 35, 37-38, 41n., 42n., 96, 97, 99, 112n., 114-116, 173, 174n., 212

Deícticos, *véase* demostrativos
Demostrativos, 54-55, 65, 76, 91, 142, 183, 184-185, 216
DEWEY, J., 153
Discurso indirecto, 34-36, 71; análisis paratáctico del, 118-120, 174-175, 183-185; forma lógica del, 108-110; teoría citativa del, 112, 116-118
Distinción analítico-sintética, 193, 195; *véase también* significados en tanto entidades
DONNE, J., 251
DUMMETT, M., 23, 114n., 123-126, 127n., 128-130, 132, 144, 179-180, 264-268

EDWARDS, W., 156
Ejecutores, 128, 130-132
ELIOT, T. S., 254, 259
Empirismo, 28, 195-199, 223
EMPSON, W., 245, 251, 261n.
Estructura profunda, 81, 144
EVANS, G., 179
Explicación teleológica, 167-168

FEYERABEND, P., 193, 196
FIELD, H., 222, 231-233, 236-238
Forma lógica, 32-38, 52-54, 87, 108-109, 160, 212-217; y estructura profunda, 81
FOSTER, J., 179, 181, 183-185
FREGE, G., 17, 37, 39-42, 44, 51, 59, 67, 69, 96, 97, 100-101, 113, 122-124, 127, 144, 204, 207, 211-212, 215, 249-250, 267
FREUD, S., 245
Fuerza, *véase* aserción; convención; imperativas
Fuerza ilocutoria, *véase* aserción; convención; imperativas

GEACH, P., 34n., 100-103, 112n., 130, 133
GOODMAN, N., 15, 246, 249n., 254
GRICE, P., 23, 153, 223

HACKING, I., 240
HARMAN, G., 23, 81n., 144, 222
Hechos, 18, 57-72, 86
HENLE, P., 245, 256
HERÁCLITO, 260
HINTIKKA, J., 120n., 129n.
Holismo, 31, 44, 150, 163, 165, 201, 205, 207, 224
HUME, D., 204, 223

Igual-decir, 71, 118
Imperativas, 123-128, 130-134
Indeterminación de traducción, 46, 79-80, 114-115, 150, 160-163, 227-228, 229, 240-241
Inescrutabilidad de la referencia, *véase* referencia, inescrutabilidad de la
Intensiones, 18, 37, 44, 89, 113-114, 121, 212, 216, 263
Interpretación radical, 137-139, 151-155, 158, 169-171, 275-276
Interrogativas, 124, 128

JACKSON, H., 108n.
JEFFREY, R., 157, 231

KANT, I., 204
KRIPKE, S., 23, 74n., 84n.
KUHN, T., 190, 192-196

LAKOFF, G., 245
LARSON, S., 23, 266
Lenguaje: en tanto aprendible, 32,74; teoría del bloque constructivo del, 28, 223; *véase también* convención, lenguaje en tanto gobernado por la; teorías de la verdad al estilo de Tarski, aplicadas a lenguajes naturales
LEWIS, D., 23, 130, 161n., 263, 273-274

MALCOLM, N., 164
MATES, B., 52, 98n., 116n.
MCCARRELL, N. S., 258
MCDOWELL, J., 179
MEAD, G. M., 153
Metáfora: en tanto ambigua, 248-249; con significado especial, 245-254, 258-260; comparada con el símil, 252; sólo con significado literal, 245, 253-256, 259-262; en tanto no parafraseable, 245
MILL, J. S., 223
Modo: distinguido de la fuerza, 124-125; análisis paratáctico del, 132-134; *véase también* imperativas; indicativas; interrogativas
MORRIS, C., 138, 223
MURRAY, J. A. H., 120n.
MURRAY, J. M., 253

OGDEN, C. K., 138, 223

ONIONS, C. T., 120n.
Ontología: relatividad de la, 230-241; revelada por una teoría de la verdad, 210-218; *véase también* referencia
Oraciones indicativas, 123-129, 132-134; *véase también* aserción
Oraciones T, 45, 47, 69, 82-83, 143-146, 148-150, 159-161, 208-210, 220-222, 225-227
Ordenes, *véase* imperativas

Palabras *versus* oraciones, 224
Parataxis, *véase* discurso indirecto; aserción; modo; imperativas; creencia, oraciones acerca de la
PARSONS, K. P., 222
Pensamiento: en tanto dependiente del lenguaje, 164-178; creencia en tanto esencial en el, 165, 177; *véase también* creencia
Pensamiento animal, *véase* actitudes proposicionales
PLATÓN, 204, 245
Primitivos semánticos, 32
Principio de caridad, *véase* caridad, principio de
PUTNAM, H., 193, 222, 237

QUINE, W. V., 14, 15, 20, 22, 23, 28-31, 32-37, 46n., 48, 54n., 79, 88n., 95-99, 101-102, 111-114, 116-117, 137n., 147n., 150, 158, 160, 162, 174, 190, 194, 196, 204, 207, 211, 223, 227, 229-230, 232-240

RAMSEY, F. P., 59, 82, 155, 157, 168
Referencia, 219-223; como explicación de la verdad, 223-228; en tanto subordinada a la verdad, 22, 229-241; inescrutabilidad de la, 22, 227-239; relatividad de la, *véase* ontología, relatividad de la
REICHENBACH, H., 99, 185
Relativismo conceptual, 22, 189-203, 235; en tanto dependiente de una distinción de esquema-contenido, 193-195, 203; en tanto reflejado en el lenguaje, 190-195
Requerimiento de finitud, 17, 32-38, 109, 113, 139, 176, 205, 210
RICHARDS, I. A., 138, 223, 245, 260
RORTY, R., 23
ROSS, J. R., 97n.
RUSSELL, B., 204
RYLE, G., 193

SAPIR, E., 195
Satisfacción, 66-68, 83-84, 142, 219-22
SCHEFFLER, I., 35-36, 116-117
SELLARS, W., 69, 116, 167
Semántica, como sintaxis recursiva más diccionario, 43; *véase también* teorías de la verdad al estilo de Tarski, en tanto teorías del significado; aplicadas a lenguajes naturales
SHAKESPEARE, W., 178, 249
Significado en tanto entidades, 39-42, 44, 88, 138
Significado: y creencia, 151-155; autonomía del, 138-139, 172, 173, 271, 272; literal, 22; *véase también* significados en tanto entidades; teorías de la verdad al estilo de Tarski, en tanto teorías del significado
SKINNER, B. F., 245
SMART, J. J. C., 193
STIEZEL, E., 120n.
STRAWSON, P., 23, 28-31, 57n., 62n., 63n., 72, 192, 204

TARSKI, A., 31-34, 45, 49-52, 57, 65-67, 68, 73, 83-90, 96-99, 101, 109, 111, 114, 123, 142-145, 149, 159-160, 180, 199, 207-208, 211, 219-221, 264
Teoría de la decisión, 155-157, 168-171
Teoría del bloque constructivo, *véase* lenguaje, teoría del bloque constructivo del
Teorías de correspondencia de la verdad, *véase* teorías de la verdad al estilo de Tarski, en tanto teorías de correspondencia; verdad, como correspodencia con los hechos
Teorías de la verdad al estilo de Tarski, 31, 45, 74, 140, 159, 175, 180-183, 219-220, 225, 230, 264; aplicadas a lenguajes naturales, 49-52, 56, 73-78, 88-91, 143-144, 199, 208-209; como teorías de correspondencia, 67-69, 87; en tanto teorías del significado, 18, 45-47, 73-75, 79-81, 93-95, 144-148, 152-163, 225-228, 232; *véase también* Convención T; Tarski; verdad; oraciones T; satisfacción
Teorías sustitucionales de la cuantificación, 19, 86

Términos evaluativos, 52-53, 143
Términos singulares, 39-40; eliminabilidad de los, 28-29
THARP, L., 84n.
Traducción, 141, 182, 190-192, 195-197, 200-202, 224; *véase también* indeterminación de traducción

VERBRUGGE, R. R., 258
Verdad: conceptos absolutos y relativizados de, 85-86, 221; como propiedad de las emisiones, 54-55, 62-65, 91, 131; como correspondencia con los hechos, 57, 61-62; teoría de redundancia de la, 58-59; *véase también* teorías de la verdad al estilo de Tarski
VERMAZEN, B., 23, 51n.

WALLACE, J., 23, 61n., 75n., 84n., 89n., 144, 227n., 231, 237n., 238
WEINREICH, U., 245
WEINSTEIN, S., 54n., 142n., 209n.
WHISTLER, J. A. M., 108
WHORF, B. L., 189-190, 195
WILDE, O., 110
WILSON, N., 20
WINTERS, I., 245
WITTGENSTEIN, L., 153, 193, 204, 261
WOOLF, V., 252

LOS CONTENIDOS DE ESTE LIBRO PUEDEN SER
REPRODUCIDOS EN TODO O EN PARTE, SIEMPRE
Y CUANDO SE CITE LA FUENTE Y SE HAGA CON
FINES ACADÉMICOS Y NO COMERCIALES

Filosofía/Lingüística
Serie CLA·DE·MA

gedisa
editorial

Donald Davidson
De la verdad y de la interpretación

¿Cuál es la relación entre el lenguaje en sus diversas modalidades y niveles –desde la pura constatación o descripción, hasta la metáfora, la cita, la manifestación de creencias y opiniones– con la realidad? ¿Hasta qué punto y con ayuda de qué reglas y leyes podemos establecer que las afirmaciones del lenguaje natural son verdaderas? ¿Qué permite al receptor de un mensaje lingüístico interpretarlo correctamente o incluso traducirlo? Estas y otras preguntas relacionadas han inquietado a la filosofía del lenguaje cuanto menos desde principios del siglo XX y en ellas se refleja, como trasfondo, el definitivo abandono de cualquier garantía metafísica de la verdad y la comprensión del lenguaje humano. Por esto, su grado de verdad e interpretabilidad ya sólo puede explorarse a partir de criterios rigurosamente orientados por la racionalidad inherente al habla, la suposición de que existe una predisposición de comunicación y comprensión y un pertinente anclaje del lenguaje en la experiencia y la percepción.

Donald Davidson, uno de los representantes contemporáneos más distinguidos de la filosofía analítica del lenguaje analiza críticamente las teorías más influyentes de su disciplina –como las de Carnap, Tarski, Chomsky y Church– en función de la construcción de sus propias tesis sobre las posibilidades y límites de probar la verdad y de comprender correctamente los enunciados del lenguaje natural. Los ensayos reunidos en este volumen se cuentan entre las contribuciones más importantes a la filosofía del lenguaje de las últimas décadas.

Donald Davidson es profesor de Filosofía en la Universidad de Berkeley (California). Por sus posiciones a menudo rebeldes frente a las corrientes teóricas predominantes, se ha convertido en una figura casi de culto para muchos jóvenes filósofos.

ISBN 84-7432-342-8
9 788474 323429
302332